

# Clase 11:

## Reestructura de bases de datos

The logo for STATA, featuring the word "STATA" in a bold, blue, sans-serif font.The logo for Python, featuring the two interlocking snakes (one blue, one yellow) followed by the word "python" in a gray, lowercase, sans-serif font.

# Contenido

1. La operación de hoy
2. Juguemos en Python
3. Juguemos en Stata
4. Juguemos en R

# 1. Recordemos la operación

The logo for Stata, featuring the word "STATA" in a bold, blue, sans-serif font. The letters are slightly stylized, with the 'S' and 'A' having unique shapes.

	departamento	frecuen~2016	frecuen~2017	frecuen~2018	frecuen~2019
1	ANTIOQUIA	19	21	30	10
2	ARAUCA	2	1	9	2
3	ATLANTICO	3	2	2	.
4	BOGOTA	1	2	3	.
5	BOLIVAR	1	1	2	5
6	BOYACA	.	.	1	.
7	CALDAS	.	4	1	.
8	CAQUETA	4	3	10	8
9	CASANARE	2	1	.	2
10	CAUCA	47	30	33	23



	departamento	año	frecuencia
1	ANTIOQUIA	2016	19
2	ANTIOQUIA	2017	21
3	ANTIOQUIA	2018	30
4	ANTIOQUIA	2019	10
5	ARAUCA	2016	2
6	ARAUCA	2017	1
7	ARAUCA	2018	9
8	ARAUCA	2019	2
9	ATLANTICO	2016	3
10	ATLANTICO	2017	2
11	ATLANTICO	2018	2
12	BOGOTA	2016	1
13	BOGOTA	2017	2
14	BOGOTA	2018	3

# ¿Cuál es long (larga)? ¿Cuál es wide (ancha)?

	departamento	frecuen~2016	frecuen~2017	frecuen~2018	frecuen~2019
1	ANTIOQUIA	19	21	30	10
2	ARAUCA	2	1	9	2
3	ATLANTICO	3	2	2	.
4	BOGOTA	1	2	3	.
5	BOLIVAR	1	1	2	5
6	BOYACA	.	.	1	.
7	CALDAS	.	4	1	.
8	CAQUETA	4	3	10	8
9	CASANARE	2	1	.	2
10	CAUCA	47	30	33	23



	departamento	año	frecuencia
1	ANTIOQUIA	2016	19
2	ANTIOQUIA	2017	21
3	ANTIOQUIA	2018	30
4	ANTIOQUIA	2019	10
5	ARAUCA	2016	2
6	ARAUCA	2017	1
7	ARAUCA	2018	9
8	ARAUCA	2019	2
9	ATLANTICO	2016	3
10	ATLANTICO	2017	2
11	ATLANTICO	2018	2
12	BOGOTA	2016	1
13	BOGOTA	2017	2
14	BOGOTA	2018	3



# ¿Cuál es long (larga)? ¿Cuál es wide (ancha)?

	departamento	frecuen~2016	frecuen~2017	frecuen~2018	frecuen~2019
1	ANTIOQUIA	19	21	30	10
2	ARAUCA	2	1	9	2
3	ATLANTICO	3	2	2	.
4	BOGOTA	1	2	3	.
5	BOLIVAR	1	1	2	5
6	BOYACA	.	.	1	.
7	CALDAS	.	4	1	.
8	CAQUETA	4	3	10	8
9	CASANARE	2	1	.	2
10	CAUCA	47	30	33	23

Wide / Ancha



	departamento	año	frecuencia
1	ANTIOQUIA	2016	19
2	ANTIOQUIA	2017	21
3	ANTIOQUIA	2018	30
4	ANTIOQUIA	2019	10
5	ARAUCA	2016	2
6	ARAUCA	2017	1
7	ARAUCA	2018	9
8	ARAUCA	2019	2
9	ATLANTICO	2016	3
10	ATLANTICO	2017	2
11	ATLANTICO	2018	2
12	BOGOTA	2016	1
13	BOGOTA	2017	2
14	BOGOTA	2018	3

Long / Larga

# ¿Cuál es la unidad de observación en cada una?

	departamento	frecuen~2016	frecuen~2017	frecuen~2018	frecuen~2019
1	ANTIOQUIA	19	21	30	10
2	ARAUCA	2	1	9	2
3	ATLANTICO	3	2	2	.
4	BOGOTA	1	2	3	.
5	BOLIVAR	1	1	2	5
6	BOYACA	.	.	1	.
7	CALDAS	.	4	1	.
8	CAQUETA	4	3	10	8
9	CASANARE	2	1	.	2
10	CAUCA	47	30	33	23

Wide / Ancha



	departamento	año	frecuencia
1	ANTIOQUIA	2016	19
2	ANTIOQUIA	2017	21
3	ANTIOQUIA	2018	30
4	ANTIOQUIA	2019	10
5	ARAUCA	2016	2
6	ARAUCA	2017	1
7	ARAUCA	2018	9
8	ARAUCA	2019	2
9	ATLANTICO	2016	3
10	ATLANTICO	2017	2
11	ATLANTICO	2018	2
12	BOGOTA	2016	1
13	BOGOTA	2017	2
14	BOGOTA	2018	3

Long / Larga

# ¿Cuál es la unidad de observación en cada una?

	departamento	frecuen~2016	frecuen~2017	frecuen~2018	frecuen~2019
1	ANTIOQUIA	19	21	30	10
2	ARAUCA	2	1	9	2
3	ATLANTICO	3	2	2	.
4	BOGOTA	1	2	3	.
5	BOLIVAR	1	1	2	5
6	BOYACA	.	.	1	.
7	CALDAS	.	4	1	.
8	CAQUETA	4	3	10	8
9	CASANARE	2	1	.	2
10	CAUCA	47	30	33	23

Departamento



	departamento	año	frecuencia
1	ANTIOQUIA	2016	19
2	ANTIOQUIA	2017	21
3	ANTIOQUIA	2018	30
4	ANTIOQUIA	2019	10
5	ARAUCA	2016	2
6	ARAUCA	2017	1
7	ARAUCA	2018	9
8	ARAUCA	2019	2
9	ATLANTICO	2016	3
10	ATLANTICO	2017	2
11	ATLANTICO	2018	2
12	BOGOTA	2016	1
13	BOGOTA	2017	2
14	BOGOTA	2018	3

Departamento en cada año



# ¿Cuáles son las variables de identificación?

	departamento	frecuen~2016	frecuen~2017	frecuen~2018	frecuen~2019
1	ANTIOQUIA	19	21	30	10
2	ARAUCA	2	1	9	2
3	ATLANTICO	3	2	2	.
4	BOGOTA	1	2	3	.
5	BOLIVAR	1	1	2	5
6	BOYACA	.	.	1	.
7	CALDAS	.	4	1	.
8	CAQUETA	4	3	10	8
9	CASANARE	2	1	.	2
10	CAUCA	47	30	33	23



	departamento	año	frecuencia
1	ANTIOQUIA	2016	19
2	ANTIOQUIA	2017	21
3	ANTIOQUIA	2018	30
4	ANTIOQUIA	2019	10
5	ARAUCA	2016	2
6	ARAUCA	2017	1
7	ARAUCA	2018	9
8	ARAUCA	2019	2
9	ATLANTICO	2016	3
10	ATLANTICO	2017	2
11	ATLANTICO	2018	2
12	BOGOTA	2016	1
13	BOGOTA	2017	2
14	BOGOTA	2018	3

## 2. Juguemos en Python

**STATA**



# Pivot

df

	foo	bar	baz	zoo
0	one	A	1	x
1	one	B	2	y
2	one	C	3	z
3	two	A	4	q
4	two	B	5	w
5	two	C	6	t



```
df.pivot(index='foo',  
          columns='bar',  
          values='baz')
```

bar	A	B	C
foo			
one	1	2	3
two	4	5	6

```
In [7]: pivoted = df.pivot(index="date", columns="variable", values="value")
```

```
In [8]: pivoted
```

```
Out[8]:
```

variable	A	B	C	D
date				
2000-01-03	0.469112	-1.135632	0.119209	-2.104569
2000-01-04	-0.282863	1.212112	-1.044236	-0.494929
2000-01-05	-1.509059	-0.173215	-0.861849	1.071804

```
In [1]: import pandas._testing as tm
```

```
In [2]: def unpivot(frame):  
...:     N, K = frame.shape  
...:     data = {  
...:         "value": frame.to_numpy().ravel("F"),  
...:         "variable": np.asarray(frame.columns).repeat(N),  
...:         "date": np.tile(np.asarray(frame.index), K),  
...:     }  
...:     return pd.DataFrame(data, columns=["date", "variable", "value"])  
...:
```

```
In [3]: df = unpivot(tm.makeTimeDataFrame(3))
```

1.Importar los datos

2.Paquete pandas

3.pivot, unpivot

### 3. Juguemos en Stata

The logo for Stata, featuring the word "STATA" in a bold, blue, sans-serif font. The letters are slightly italicized and have a modern, clean design.

<i>long</i>						<i>wide</i>		
<i>i</i>	<i>j</i>	<i>stub</i>				<i>i</i>	<i>stub1</i>	<i>stub2</i>
1	1	4.1	← reshape →			1	4.1	4.5
1	2	4.5				2	3.3	3.0
2	1	3.3						
2	2	3.0						

To go from long to wide:

`reshape wide stub, i(i) j(j)` *j* existing variable

To go from wide to long:

`reshape long stub, i(i) j(j)` *j* new variable

1.Cargar los datos

2.reshape wide

3.reshape long

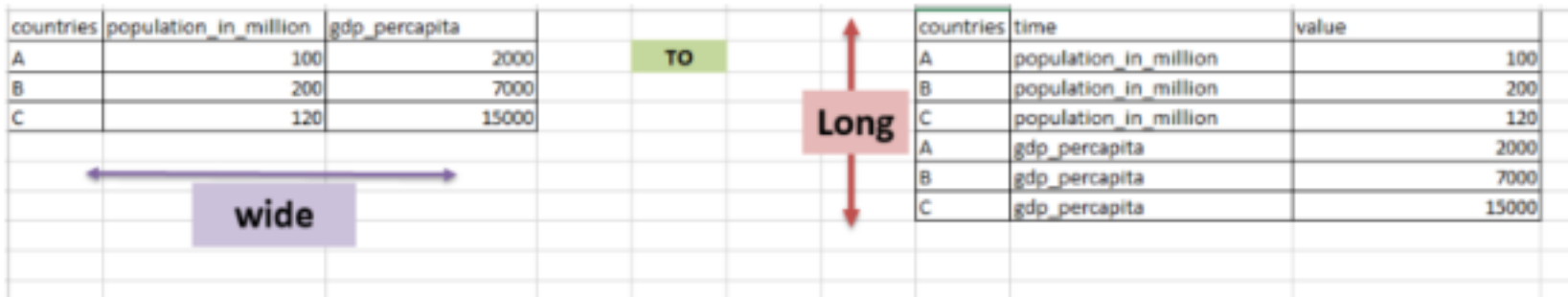


## 4. Juguemos en R

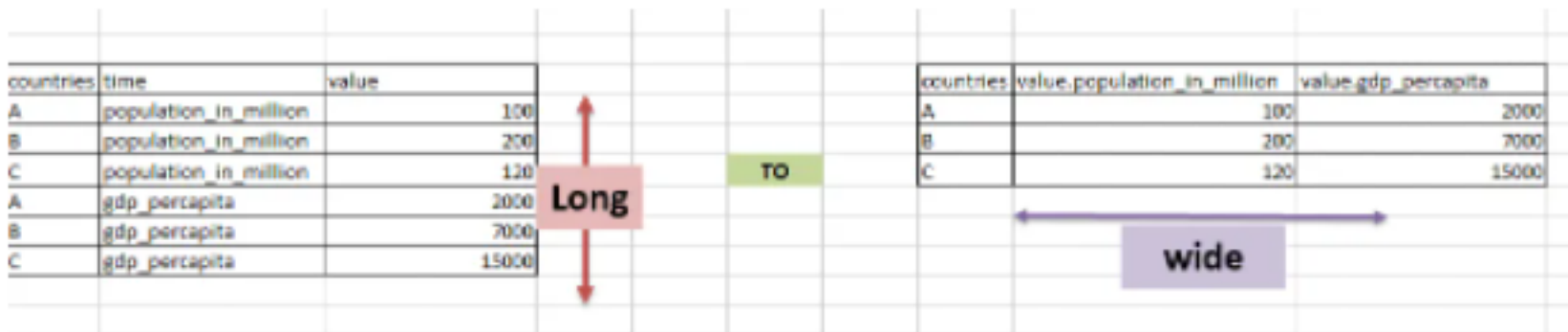
**STATA**



 python



```
country_w_to_L<- reshape(data=country, idvar="countries",
                          varying = c("population_in_million","gdp_percapita"),
                          v.name=c("value"),
                          times=c("population_in_million","gdp_percapita"),
                          new.row.names = 1:1000,
                          direction="long")
```



```
country_L_to_w <- reshape(data=country_w_to_L, idvar="countries",  
                           v.names = "value",  
                           timevar = "time",  
                           direction="wide")
```

1.Cargar los datos

2.reshape()

# Referencias

- [https://pandas.pydata.org/docs/user\\_guide/reshaping.html](https://pandas.pydata.org/docs/user_guide/reshaping.html)
- <https://www.stata.com/manuals13/dreshape.pdf>
- <https://www.datasciencemadesimple.com/reshape-in-r-from-wide-to-long-from-long-to-wide/#:~:text=Reshape%20function%20in%20R%20transforms,another%20way%20of%20transforming%20data.>