

Foundations of R Software

Lecture 44

Data Frames: Combining and Merging

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Data Frames: Combining

□ There are three main techniques :

`cbind()` – combining the columns of two data frames side-by-side.

`merge()` – joining two data frames using a common column.

`rbind()` – stacking two data frames on top of each other,
appending one to the other.

Data Frames: Combining

- ❑ The command **`cbind`** horizontally merges two data frames side by side.

Example: Create two data frames as follows:

```
df1=data.frame(state=c("UP", "MP", "AP", "JK"),  
               popnsize=c(1000,2000,3000,4000))
```

```
df2=data.frame(state=c("UP", "MP", "AP", "JK"),  
               samplesize=c(100,200,300,400),  
               surveycompleted=c("Yes", "No", "Yes", "No"))
```

Data Frames: Combining

```
> df1
```

	state	popnsize
1	UP	1000
2	MP	2000
3	AP	3000
4	JK	4000

```
> df2
```

	state	samplesize	surveycompleted
1	UP	100	Yes
2	MP	200	No
3	AP	300	Yes
4	JK	400	No

Data Frames: Combining

```
> cbind(df1, df2)
```

	state	popnsize	state	samplesize	surveycompleted
1	UP	1000	UP	100	Yes
2	MP	2000	MP	200	No
3	AP	3000	AP	300	Yes
4	JK	4000	JK	400	No

Data Frames: Combining

```
R Console
> df1
  state popnsize
1    UP    1000
2    MP    2000
3    AP    3000
4    JK    4000
> df2
  state samplesize surveycompleted
1    UP         100           Yes
2    MP         200           No
3    AP         300           Yes
4    JK         400           No
> cbind(df1,df2)
  state popnsize state samplesize surveycompleted
1    UP    1000    UP         100           Yes
2    MP    2000    MP         200           No
3    AP    3000    AP         300           Yes
4    JK    4000    JK         400           No
~ |
```

Data Frames: Merging

- ❑ The command **merge** horizontally merges two data frames by common columns or row names.

Example: Create two data frames as follows:

```
df1=data.frame(state=c("UP", "MP", "AP", "JK"),  
               popnsize=c(1000,2000,3000,4000))
```

```
df2=data.frame(state=c("UP", "MP", "AP", "JK"),  
               samplesize=c(100,200,300,400),  
               surveycompleted=c("Yes", "No", "Yes", "No"))
```

Variable “**state**” is common between the two data frames and we want to merge the two data frames with respect to **state**.

Data Frames: Merging

- ❑ The command **merge** horizontally merges two data frames by common columns or row names.

Usage : **merge(x, y, ...)**

Arguments :

x, y : data frames, or objects to be coerced to one.

by, by.x, by.y : specifications of the columns used for merging.

sort : logical.

no.dups : logical indicating that suffixes are appended in more cases to avoid duplicated column names in the result.

Data Frames: Merging

```
> df1
```

	state	popnsize
1	UP	1000
2	MP	2000
3	AP	3000
4	JK	4000

```
> df2
```

	state	samplesize	surveycompleted
1	UP	100	Yes
2	MP	200	No
3	AP	300	Yes
4	JK	400	No

```
> merge(df1, df2, by="state")
```

	state	popnsize	samplesize	surveycompleted
1	AP	3000	300	Yes
2	JK	4000	400	No
3	MP	2000	200	No
4	UP	1000	100	Yes

Data Frames: Merging

```
R Console
> df1
  state popnsiz
1    UP   1000
2    MP   2000
3    AP   3000
4    JK   4000
>
> df2
  state samplesize surveycompleted
1    UP         100             Yes
2    MP         200             No
3    AP         300             Yes
4    JK         400             No
> merge(df1,df2,by="state")
  state popnsiz samplesize surveycompleted
1    AP   3000         300             Yes
2    JK   4000         400             No
3    MP   2000         200             No
4    UP   1000         100             Yes
> |
```

Data Frames: Combining vertically

- ❑ The command `rbind` stacks two data frames on top of each other, appending one to the other

Example: Create two data frames as follows:

```
df11=data.frame(state=c("UP", "MP", "AP",  
    "JK"), popnsize=c(1000,2000,3000,4000))
```

```
df22=data.frame(state=c("Bihar", "Delhi",  
    "Punjab"), popnsize =c(100,200,300))
```

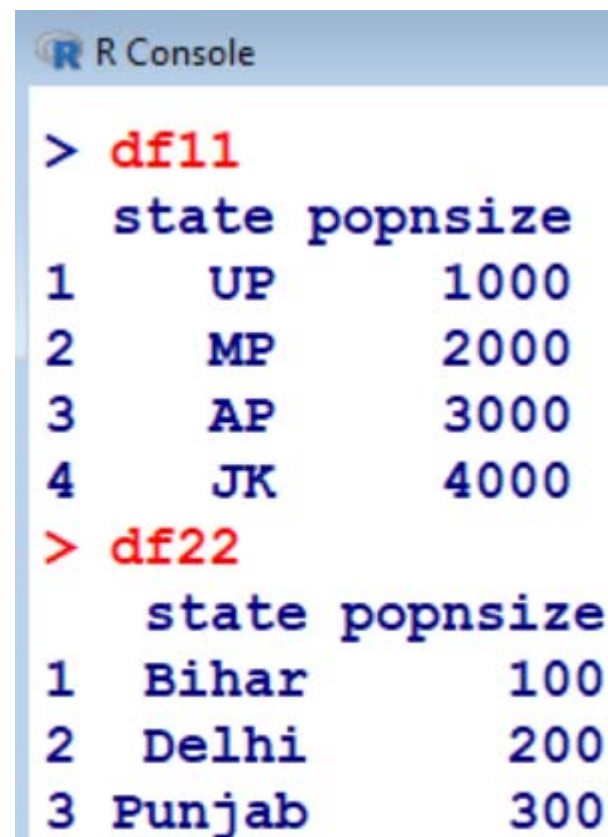
Data Frames: Combining vertically

> df11

	state	popnsize
1	UP	1000
2	MP	2000
3	AP	3000
4	JK	4000

> df22

	state	popnsize
1	Bihar	100
2	Delhi	200
3	Punjab	300



```
R Console
> df11
  state popnsize
1    UP    1000
2    MP    2000
3    AP    3000
4    JK    4000
> df22
  state popnsize
1 Bihar     100
2 Delhi     200
3 Punjab     300
```

Data Frames: Combining vertically

```
> rbind(df11, df22)
```

	state	popnsize
1	UP	1000
2	MP	2000
3	AP	3000
4	JK	4000
5	Bihar	100
6	Delhi	200
7	Punjab	300

R Console

```
> df11
  state popnsize
1    UP    1000
2    MP    2000
3    AP    3000
4    JK    4000
> df22
  state popnsize
1 Bihar     100
2 Delhi     200
3 Punjab     300
> rbind(df11, df22)
  state popnsize
1    UP    1000
2    MP    2000
3    AP    3000
4    JK    4000
5 Bihar     100
6 Delhi     200
7 Punjab     300
```