

DATA SCIENCE WITH R

Class 4 – Data Manipulation in R

Topic 4

★ Merging Tables ★

INDEX

Manipulating data using base R

Using dplyr to manipulate data

Working with date objects



Merging tables

Missing value treatment

Using reshape2() to transpose data

Manipulating Character Strings

Using sqldf

Joining dataframes

Joining dataframes

- Just like tables can be joined in sql, we can perform joins on dataframes in R
- Following types of joins can be accomplished
- Inner Join
- Left outer join
- Right outer join
- Full outer join

Joining dataframes

- Inner join: Joining two tables based on a key column, such that rows matching in both tables are selected

	CustomerId	Product		CustomerId	State
1	1	Toaster	1	2	Alabama
2	2	Toaster	2	4	Alabama
3	3	Toaster	3	6	Ohio
4	4	Radio			
5	5	Radio			
6	6	Radio			

Joining dataframes

- Inner join: Joining two tables based on a key column, such that rows matching in both tables are selected

	CustomerId	Product
1	1	Toaster
2	2	Toaster
3	3	Toaster
4	4	Radio
5	5	Radio
6	6	Radio

	CustomerId	State
1	2	Alabama
2	4	Alabama
3	6	Ohio

```
> merge(x=df1,y=df2,by="CustomerId")#Inner Join/Intersection of both tables
```

	CustomerId	Product	State
1	2	Toaster	Alabama
2	4	Radio	Alabama
3	6	Radio	Ohio

Joining dataframes

- Full Outer Join: Two tables are joined irrespective of any match between the rows

	CustomerId	Product
1	1	Toaster
2	2	Toaster
3	3	Toaster
4	4	Radio
5	5	Radio
6	6	Radio

	CustomerId	State
1	2	Alabama
2	4	Alabama
3	6	Ohio

```
> merge(x = df1, y = df2, by = "CustomerId", all = TRUE)#Outer join:
```

	CustomerId	Product	State
1	1	Toaster	<NA>
2	2	Toaster	Alabama
3	3	Toaster	<NA>
4	4	Radio	Alabama
5	5	Radio	<NA>
6	6	Radio	Ohio

Joining dataframes

- Left Outer Join: All the rows of left table are retained while matching rows of right table are displayed

	CustomerId	Product		CustomerId	State
1	1	Toaster	1	2	Alabama
2	2	Toaster	2	4	Alabama
3	3	Toaster	3	6	ohio
4	4	Radio			
5	5	Radio			
6	6	Radio			

```
> merge(x = df1, y = df2, by = "CustomerId", all.x=TRUE)#Left join
```

	CustomerId	Product	State
1	1	Toaster	<NA>
2	2	Toaster	Alabama
3	3	Toaster	<NA>
4	4	Radio	Alabama
5	5	Radio	<NA>
6	6	Radio	ohio

Joining dataframes

- Right Outer Join: All the rows of right table are retained while matching rows of left table are displayed

	CustomerId	Product
1	1	Toaster
2	2	Toaster
3	3	Toaster
4	4	Radio
5	5	Radio
6	6	Radio

	CustomerId	State
1	2	Alabama
2	4	Alabama
3	6	Ohio

```
> merge(x = df1, y = df2, by = "CustomerId", all.y=TRUE)#Right join
```

	CustomerId	Product	State
1	2	Toaster	Alabama
2	4	Radio	Alabama
3	6	Radio	Ohio

RECAP

- Inner Join
- Left outer join
- Right outer join
- Full outer join