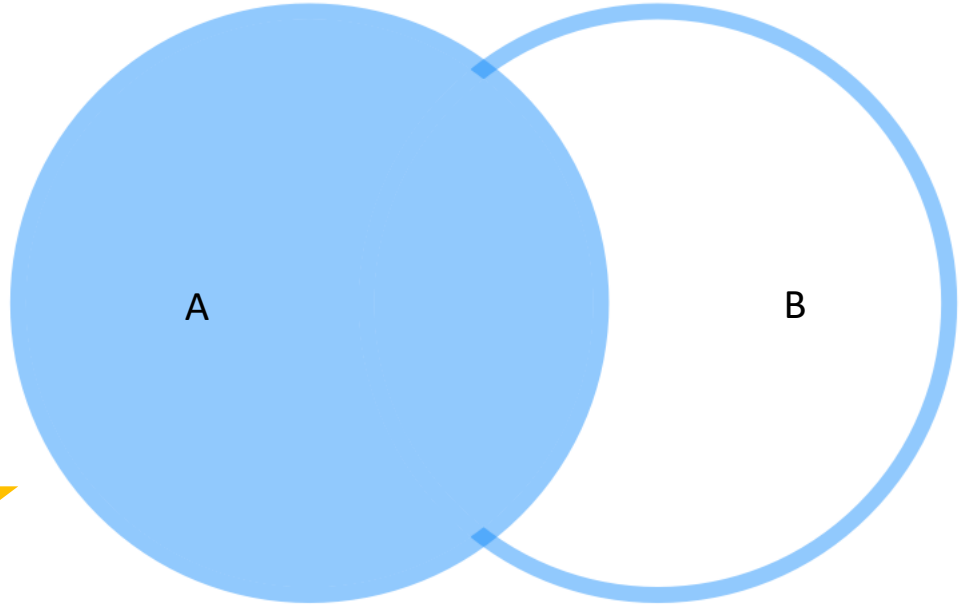
Left and right outer joins

Merging: left outer join



Left outer joins keep all data in set A and the matching data in set B (1)

```
merge(A, B, by="ID", all.x=TRUE)
```

Common identifier (2)

Use all observations from A, but only the matching from B (3)

Left join merges on all of the common identifiers in the left specified table (1/2)

A (myData)

Customer	TransDate	Quantity	PurchAmount	Cost
149332	2005-11-15	1	199.95	107.00
172951	2008-08-29	1	199.95	108.00
120621	2007-10-19	1	99.95	49.00
149236	2005-11-14	1	39.95	18.95
149236	2007-12-06	1	79.95	35.00
…	…	…	…	…

B (CustData)

Customer	Gender	Birthdate	ZIP	JoinDate
80365	f	1991-08-26	US-06332	2009-09-15
149332	m	1998-07-07	US-08873	2005-11-05
84374	m	1977-07-10	US-06400	1988-08-10
149236	f	1955-08-15	US-92646	1971-02-16
100001	m	1974-05-08	US-02332	1992-02-21
…	…	…	…	…

```
merge(myData, CustData, by="Customer", all.x=TRUE)
```

Left join merges on all of the common identifiers in the left specified table (2/2)

A (myData)

Customer	TransDate	Quantity	PurchAmount	Cost
149332	2005-11-15	1	199.95	107.00
172951	2008-08-29	1	199.95	108.00
120621	2007-10-19	1	99.95	49.00
149236	2005-11-14	1	39.95	18.95
149236	2007-12-06	1	79.95	35.00
...	...	...	...	...

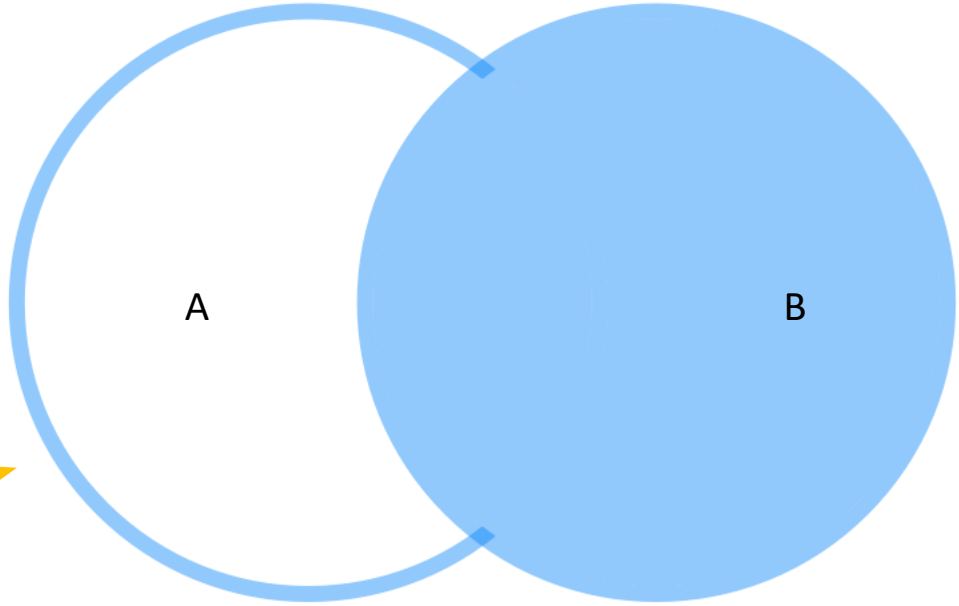
B (CustData)

Customer	Gender	Birthdate	ZIP	JoinDate
80365	f	1991-08-26	US-06332	2009-09-15
149332	m	1998-07-07	US-08873	2005-11-05
84374	m	1977-07-10	US-06400	1988-08-10
149236	f	1955-08-15	US-92646	1971-02-16
100001	m	1974-05-08	US-02332	1992-02-21
...	...	...	...	...

Append column entries of table B if customer ID matches with customer ID of table A.

Customer	TransDate	Quantity	PurchAmount	Cost	Gender	Birthdate	ZIP	JoinDate
149332	2005-11-15	1	199.95	107.00	m	1998-07-07	US-08873	2005-11-05
172951	2008-08-29	1	199.95	108.00	NA	NA	NA	NA
120621	2007-10-19	1	99.95	49.00	NA	NA	NA	NA
149236	2005-11-14	1	39.95	18.95	m	1998-07-07	US-08873	2005-11-05
149236	2001-06-12	1	79.95	35.00	m	1998-07-07	US-08873	2005-11-05
...	...	...	...	...	...	...	...	...

Left joins are the same as right joins but the other way around



1. Right outer joins keep all data in set B and the matching data in set A

```
merge(A, B, by="ID", all.y=TRUE)
```

2. Common identifier

3. Use all observations from B, but only the matching from A

Right join merges on all of the common identifiers in the table specified on the right side

A (myData)

Customer	TransDate	Quantity	PurchAmount	Cost
149332	2005-11-15	1	199.95	107.00
172951	2008-08-29	1	199.95	108.00
120621	2007-10-19	1	99.95	49.00
149236	2005-11-14	1	39.95	18.95
149236	2007-12-06	1	79.95	35.00
…	…	…	…	…

B (CustData)

Customer	Gender	Birthdate	ZIP	JoinDate
80365	f	1991-08-26	US-06332	2009-09-15
149332	m	1998-07-07	US-08873	2005-11-05
84374	m	1977-07-10	US-06400	1988-08-10
149236	f	1955-08-15	US-92646	1971-02-16
100001	m	1974-05-08	US-02332	1992-02-21
…	…	…	…	…

```
merge(myData, CustData, by="Customer", all.y=TRUE)
```

Right join merges on all of the common identifiers in the table specified on the right side

A (`myData`)

Customer	TransDate	Quantity	PurchAmount	Cost
149332	2005-11-15	1	199.95	107.00
172951	2008-08-29	1	199.95	108.00
120621	2007-10-19	1	99.95	49.00
149236	2005-11-14	1	39.95	18.95
149236	2007-12-06	1	79.95	35.00
...	...	...	...	...

B (`CustData`)

Customer	Gender	Birthdate	ZIP	JoinDate
80365	f	1991-08-26	US-06332	2009-09-15
149332	m	1998-07-07	US-08873	2005-11-05
84374	m	1977-07-10	US-06400	1988-08-10
149236	f	1955-08-15	US-92646	1971-02-16
100001	m	1974-05-08	US-02332	1992-02-21
...	...	...	...	...

Append column entries of table A if customer ID matches with customer ID of table B.

Customer	TransDate	Quantity	PurchAmount	Cost	Gender	Birthdate	ZIP	JoinDate
80365	NA	NA	NA	NA	f	1991-08-26	US-06332	2009-09-15
149332	2005-11-15	1	199.95	107.00	m	1987-05-04	US-08873	2005-11-05
84374	NA	NA	NA	NA	m	1977-07-10	US-06400	1988-08-10
149236	2005-11-14	1	39.95	18.95	f	1955-08-15	US-92646	1971-02-16
100001	NA	NA	NA	NA	m	1974-05-08	US-02332	1992-02-21
...	...	...	...	...	...	...	...	...

When do we need left and right joins?

- Left and right joins return all the data available in the left or right data.table.
- For example:
  - We want all the data for all customers who made a transaction (left outer).
  - We want all the data for all the customers in our data.table (right outer).