

Week 8: Combining Data Sets with `dplyr`

Suppose you have the following two data sets. The first, `df1` has the variables `id_num` and `xvar`. The second, `df2` has the variables `id` and `yvar`. `id_num` and `id` serve as identification variables, possibly with duplicates, where observations from the first data set with `id_num = 1` correspond to observations in the second data set with `id = 1`.

id_num	xvar
1	16
1	-1
2	11
4	13

id	yvar
1	-1
2	-4
2	0
3	-9

Mutating Joins

- `left_join()`

- `right_join()`

- `inner_join()`

id_numb	xvar
1	16
1	-1
2	11
4	13

id	yvar
1	-1
2	-4
2	0
3	-9

- `full_join()`

Filtering Joins

- `semi_join()`

- `anti_join()`