Joining Data Sets with dplyr

Dan Tofan
SOFTWARE ENGINEER, PHD

@dan_tofan www.programmingwithdan.com

Overview

Why dplyr?

Inner joins

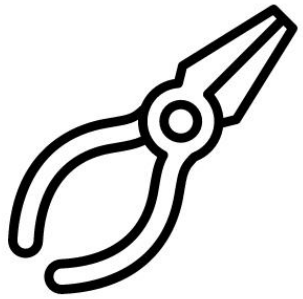
Left, right, and full outer joins

Semi joins

Anti joins

Summary

Why dplyr?

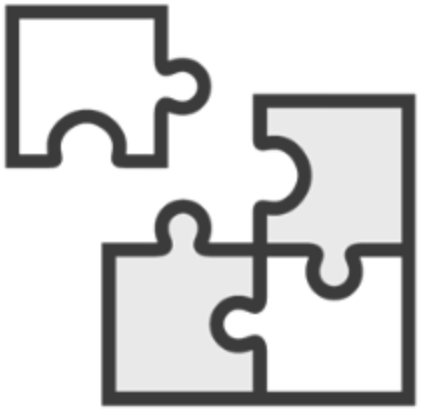


Data frames

Manipulation

Performance

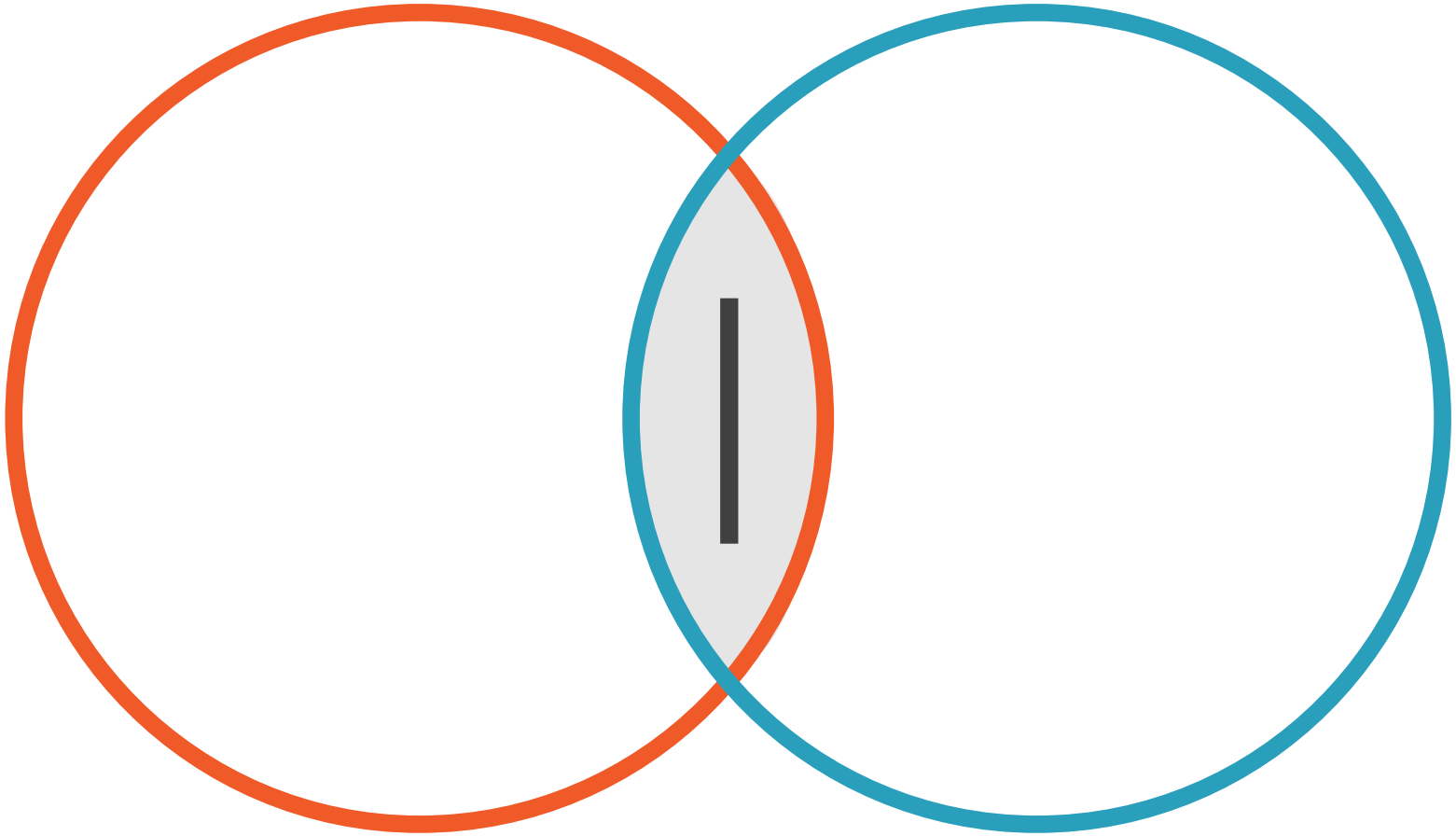
Fast!



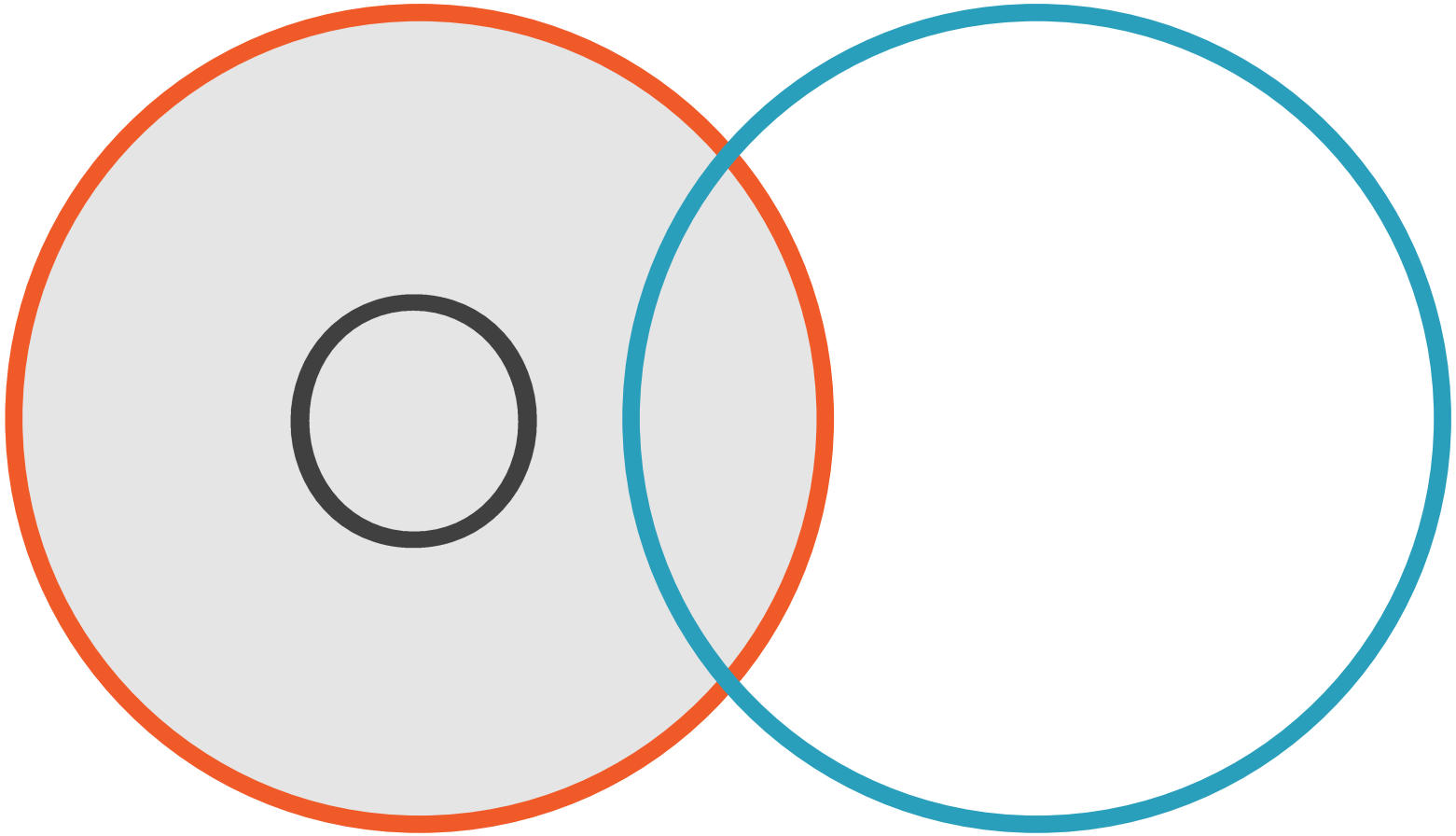
Tidyverse

Integration

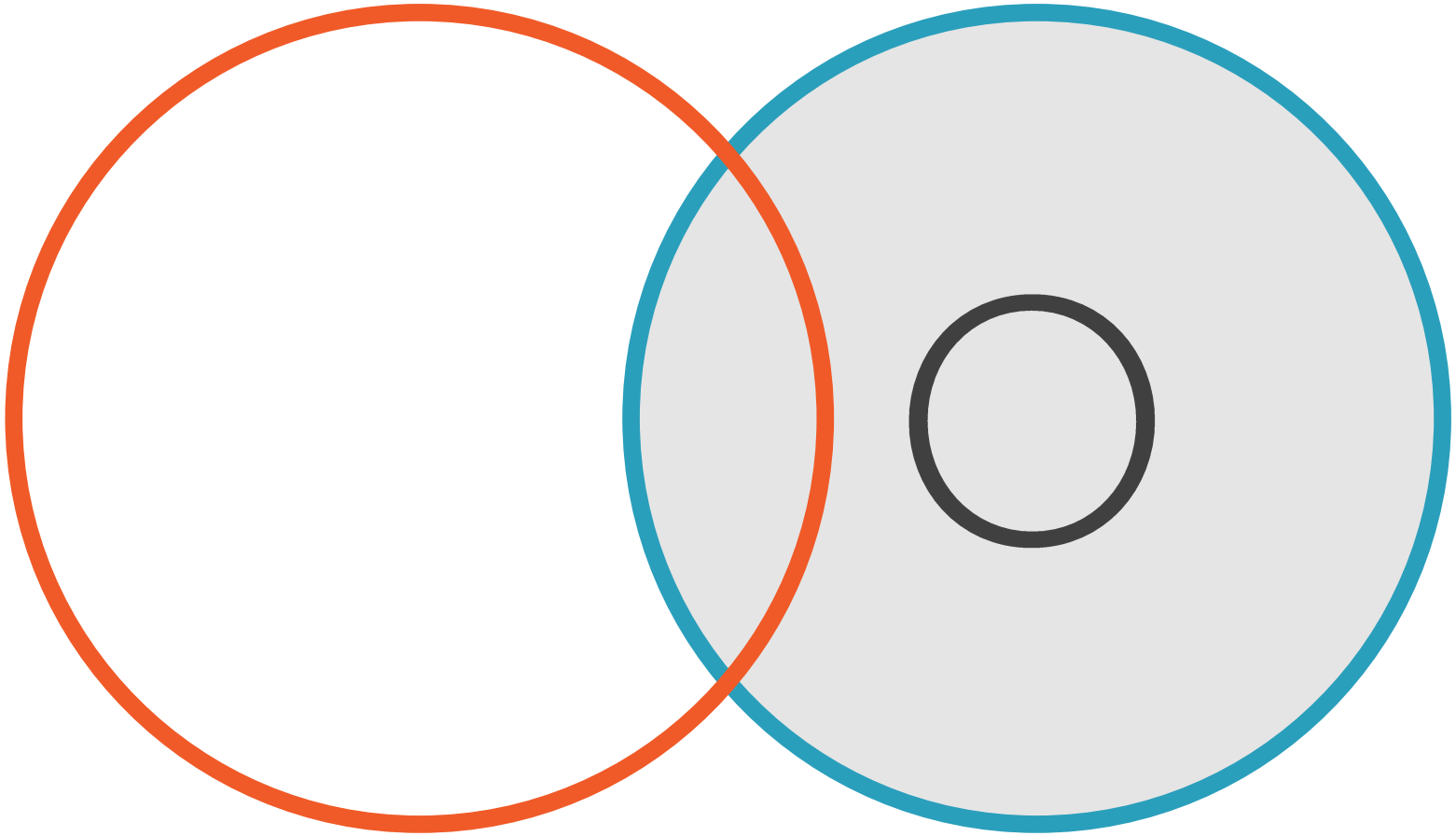
Inner Join



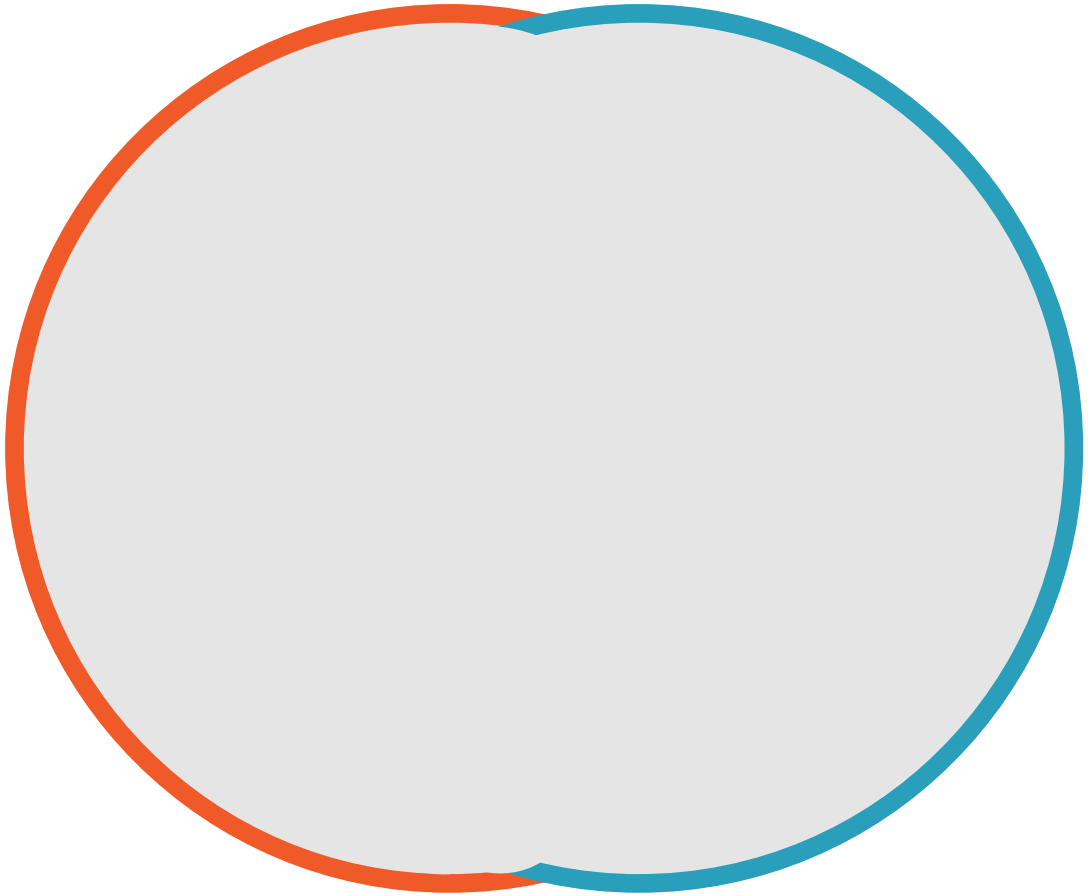
Left Outer Join



Right Outer Join



Full Join



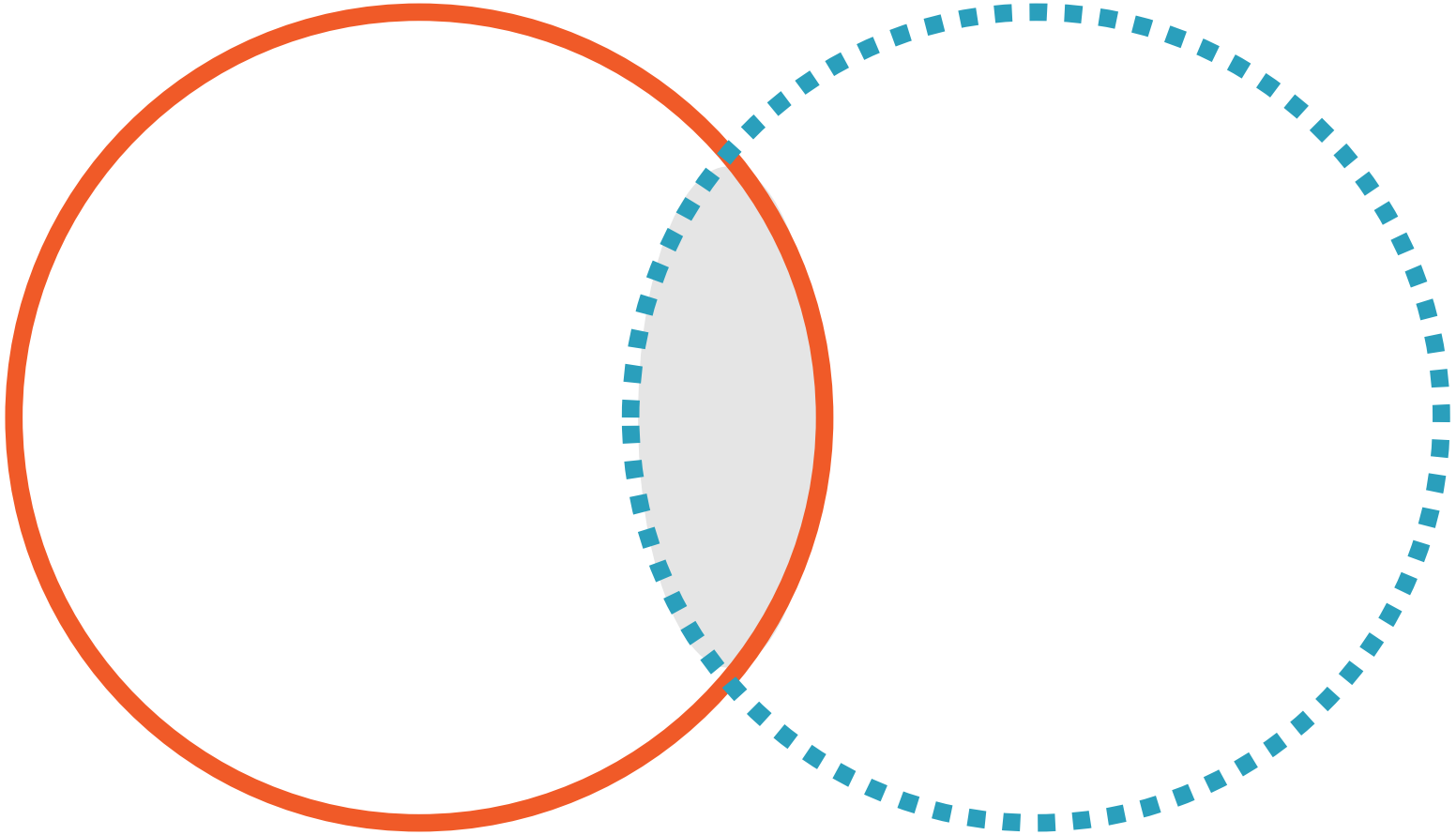
Semi Join

Customer_id	Name
1	Anne
2	John
3	Tom

Invoice_id	Total	Customer_id
1	100	1
2	200	2
3	300	2

Customer_id	Name
1	Anne
2	John

Semi Join



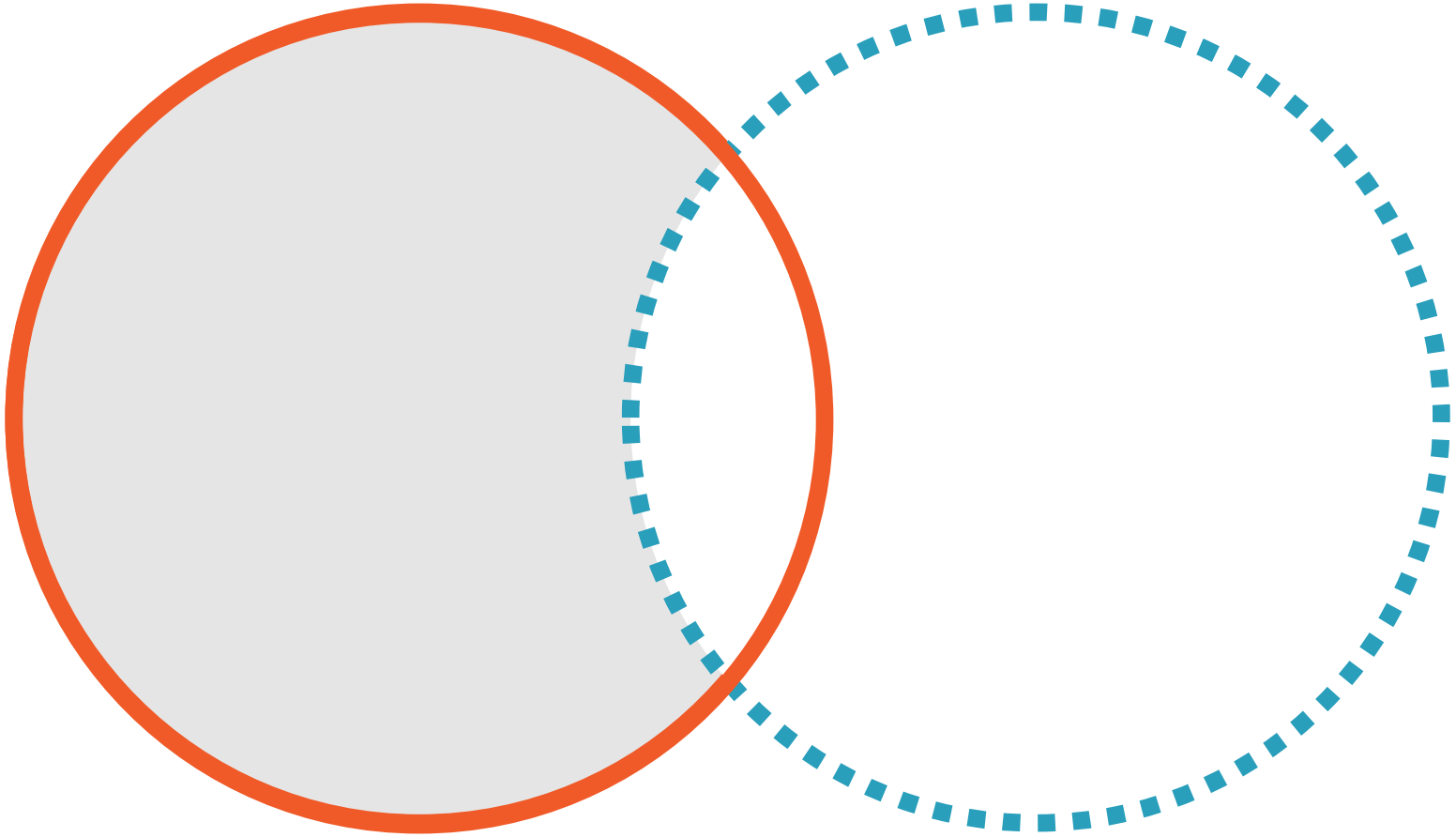
Anti Join

Customer_id	Name
1	Anne
2	John
3	Tom
4	Lucy

Invoice_id	Total	Customer_id
1	100	1
2	200	2
3	300	2

Customer_id	Name
3	Tom
4	Lucy

Anti Join



Module Summary

Why dplyr?

Inner joins

Left, right, and full outer joins

Semi joins

Anti joins

Course Summary

Merging vectors

Joining data sets with the merge() function

Joining data sets with dplyr

What is Next?

Expand course examples

Use your own data

Explore dplyr