

Trashing the Database

How R does SQL
IRE 2018

tinyurl.com/IRE2018R

JOINING DATA

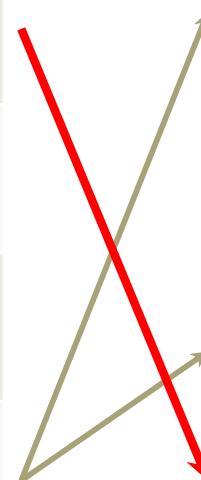
- Link tables and databases
- Why databases are databases
- Two main kinds
 - Inner Join
 - Outer Join



INNER JOINS

ID	DRINK
2	BOMBAY SAPPHIRE GIN MARTINI
3	JACK DANIEL NEAT
4	LAPHROAIG 10
5	BUD LIGHT

ORDER	CUSTOMER
5	T's BROTHER
NA	T's HIGH SCHOOL FRIEND
5	T's NEIGHBOR PAUL
2	T



RESULT

ID	DRINK	CUSTOMER
5	BUD LIGHT	T'S BROTHER
5	BUD LIGHT	T'S NEIGHBOR PAUL
2	BOMBAY SAPPHIRE GIN MARTINI	T

OTHER JOINS

ID	DRINK
2	BOMBAY SAPPHIRE GIN MARTINI
3	JACK DANIEL NEAT
4	LAPHROAIG 10
5	BUD LIGHT

ORDER	PERSON
5	T's BROTHER
NULL	T's HIGH SCHOOL FRIEND
5	T's NEIGHBOR
2	T

@txtianmiller

LEFT JOIN

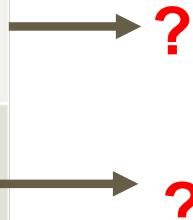
ID	DRINK	CUSTOMER
2	BOMBAY SAPPHIRE MARTINI	T
3	JACK DANIELS NEAT	NULL
4	LAPHROAIG 10	NULL
5	BUD LIGHT	T'S BROTHER
5	BUD LIGHT	T'S NEIGHBOR PAUL

OUTER JOIN

ID	DRINK	CUSTOMER
2	BOMBAY SAPPHIRE MARTINI	T
3	JACK DANIELS NEAT	NULL
4	LAPHROAIG 10	NULL
NULL	NULL	T'S HIGH SCHOOL FRIEND
5	BUD LIGHT	T'S BROTHER
5	BUD LIGHT	T'S NEIGHBOR PAUL

ANTI JOIN

ID	DRINK
2	BOMBAY SAPPHIRE GIN MARTINI
3	JACK DANIEL NEAT
4	LAPHROAIG 10
5	BUD LIGHT



ORDER	PERSON
5	T's BROTHER
NULL	T's HIGH SCHOOL FRIEND
5	T's NEIGHBOR
2	T

@txtianmiller

Useful in data cleaning

ID	DRINK	CUSTOMER
3	JACK DANIELS NEAT	NULL
4	LAPHROAIG 10	NULL

joins!

```
gator.bites <- AlligatorBites  
  
gator.bites.tot <- gator.bites %>%  
  select(water.body) %>%  
  group_by(water.body) %>%  
  summarize(bite.count = n()) %>%  
  arrange(desc(bite.count))
```

joins!

```
gator.danger <- gator.hunt.tot %>%
  inner_join(gator.bites.tot, by =
c("area.name" = "water.body")) %>%
arrange(desc(count))
```

Results

	area.name	count	bite.count
1	Lake Okeechobee - Okeechobee	14005	4
2	Lake Tarpon	21	3
3	Blue Springs Lake	2	2