

Hands on Labs: SQL JOINS

These ungraded hands-on-lab will help you build your skills working with combining information from multiple tables. Solutions to all problems are immediately available, but you'll learn the material more *durably* if you attempt each problem before consulting its solution.

All of the queries use tables in the flights database. You may find the data descriptions here helpful:

<http://cran.r-project.org/web/packages/nycflights13/nycflights13.pdf>

1. [Warm-up Exercise] Write a SELECT statement that shows for all of the flights during the period February 14th to February 17th for each of the four planes above: 'N125UW', 'N848MQ', 'N328AA', 'N247JB'. Your select statement should return the following information: tailnum, flight date information, departure delay, arrival delay, 3 digit destination code

	tailnum character(6)	year integer	month integer	day integer	dep_delay integer	arr_delay integer	dest character(3)
1	N247JB	2013	2	14	124	127	CLT
2	N247JB	2013	2	14	99	91	SJU
3	N328AA	2013	2	14	1	-16	LAX
4	N328AA	2013	2	14	12	-9	LAX
5	N328AA	2013	2	15	33	8	LAX
6	N328AA	2013	2	16	8	-35	LAX
7	N328AA	2013	2	17	-7	-19	LAX
8	N848MQ	2013	2	17	39	30	RDU
9	N848MQ	2013	2	15	-8	-7	DCA
10	N848MQ	2013	2	16	2	-16	DCA
11	N848MQ	2013	2	14	24	16	DCA
12	N848MQ	2013	2	17	-7	-22	RDU
13	N848MQ	2013	2	14	12	0	RDU
14	N848MQ	2013	2	15	-4	-16	DCA

2. Add a join clause, to also show destination airport name. How many rows were returned?

	tailnum character(6)	year integer	month integer	day integer	dep_delay integer	arr_delay integer	dest character(3)	name character varying
1	N247JB	2013	2	14	124	127	CLT	Charlotte Douglas Intl
2	N328AA	2013	2	14	12	-9	LAX	Los Angeles Intl
3	N328AA	2013	2	14	1	-16	LAX	Los Angeles Intl
4	N328AA	2013	2	15	33	8	LAX	Los Angeles Intl
5	N328AA	2013	2	16	8	-35	LAX	Los Angeles Intl
6	N328AA	2013	2	17	-7	-19	LAX	Los Angeles Intl
7	N848MQ	2013	2	14	24	16	DCA	Ronald Reagan Washington Natl
8	N848MQ	2013	2	14	12	0	RDU	Raleigh Durham Intl
9	N848MQ	2013	2	17	-7	-22	RDU	Raleigh Durham Intl
10	N848MQ	2013	2	15	-4	-16	DCA	Ronald Reagan Washington Natl
11	N848MQ	2013	2	15	-8	-7	DCA	Ronald Reagan Washington Natl
12	N848MQ	2013	2	16	2	-16	DCA	Ronald Reagan Washington Natl
13	N848MQ	2013	2	17	39	30	RDU	Raleigh Durham Intl

3. Rewrite the join as a left join, and compare your results.

	tailnum character(6)	year integer	month integer	day integer	dep_delay integer	arr_delay integer	dest character(3)	name character varying
1	N247JB	2013	2	14	124	127	CLT	Charlotte Douglas Intl
2	N247JB	2013	2	14	99	91	SJU	
3	N328AA	2013	2	14	1	-16	LAX	Los Angeles Intl
4	N328AA	2013	2	14	12	-9	LAX	Los Angeles Intl
5	N328AA	2013	2	15	33	8	LAX	Los Angeles Intl
6	N328AA	2013	2	16	8	-35	LAX	Los Angeles Intl
7	N328AA	2013	2	17	-7	-19	LAX	Los Angeles Intl
8	N848MQ	2013	2	17	39	30	RDU	Raleigh Durham Intl
9	N848MQ	2013	2	15	-8	-7	DCA	Ronald Reagan Washington Natl
10	N848MQ	2013	2	16	2	-16	DCA	Ronald Reagan Washington Natl
11	N848MQ	2013	2	14	24	16	DCA	Ronald Reagan Washington Natl
12	N848MQ	2013	2	17	-7	-22	RDU	Raleigh Durham Intl
13	N848MQ	2013	2	14	12	0	RDU	Raleigh Durham Intl
14	N848MQ	2013	2	15	-4	-16	DCA	Ronald Reagan Washington Natl

4. Add a second left join clause to also show the number of seats in the planes

	tailnum character(6)	year integer	month integer	day integer	dep_delay integer	arr_delay integer	dest character(3)	name character varying	seats integer
1	N247JB	2013	2	14	124	127	CLT	Charlotte Douglas Intl	20
2	N247JB	2013	2	14	99	91	SJU		20
3	N328AA	2013	2	14	1	-16	LAX	Los Angeles Intl	255
4	N328AA	2013	2	14	12	-9	LAX	Los Angeles Intl	255
5	N328AA	2013	2	15	33	8	LAX	Los Angeles Intl	255
6	N328AA	2013	2	16	8	-35	LAX	Los Angeles Intl	255
7	N328AA	2013	2	17	-7	-19	LAX	Los Angeles Intl	255
8	N848MQ	2013	2	17	39	30	RDU	Raleigh Durham Intl	
9	N848MQ	2013	2	15	-8	-7	DCA	Ronald Reagan Washington Natl	
10	N848MQ	2013	2	16	2	-16	DCA	Ronald Reagan Washington Natl	
11	N848MQ	2013	2	14	24	16	DCA	Ronald Reagan Washington Natl	
12	N848MQ	2013	2	17	-7	-22	RDU	Raleigh Durham Intl	
13	N848MQ	2013	2	14	12	0	RDU	Raleigh Durham Intl	
14	N848MQ	2013	2	15	-4	-16	DCA	Ronald Reagan Washington Natl	

5. Rewrite your query, using table aliases. (Results unchanged from above). Do you find the code more readable with or without table aliases?

6. What are the names of the five airports that receive the most flights?

	name character varying	count bigint
1	Chicago Ohare Intl	17283
2	Hartsfield Jackson Atlanta Intl	17215
3	Los Angeles Intl	16174
4	General Edward Lawrence Logan Intl	15508
5	Orlando Intl	14082

7. What are American Airlines' destination cities from the New York airports?

	airport character varying
1	Austin Bergstrom Intl
2	Chicago Ohare Intl
3	Dallas Fort Worth Intl
4	Eagle Co Rgnl
5	Fort Lauderdale Hollywood Intl
6	General Edward Lawrence Logan Intl
7	George Bush Intercontinental
8	Lambert St Louis Intl
9	Los Angeles Intl
10	Mc Carran Intl
11	Miami Intl
12	Orlando Intl
13	Palm Beach Intl
14	San Diego Intl
15	San Francisco Intl
16	Seattle Tacoma Intl
17	Tampa Intl