STAT 542 HW 6 Example Solutions

Problem 3, Chapter 15 exercises:

> sqldf("SELECT

+ COUNT(\*) AS N

+ FROM flights

+ WHERE year = 2013 AND month = 5 AND day = 4

+ AND dest = 'DFW'

+ LIMIT 0, 6

+ ")

N

1 15

There are **15** flights that flew into DFW on May 4, 2013.

Problem 5 in the Chapter 15 Exercises (in Section 15.8 of the book).

The ten most common destinations from JFK in 2013, as well as the number of flights in 2013 for each destination, are:

> sqldf("SELECT

+ dest,

+ COUNT(\*) AS numFlights

+ FROM flights

+ WHERE year = 2013 AND origin = 'JFK'

+ GROUP BY dest

+ ORDER BY numFlights DESC

+ LIMIT 0, 10

+ ")

dest numFlights

1 LAX 11262

2 SFO 8204

3 BOS 5898

4 MCO 5464

5 SJU 4752

6 FLL 4254

7 LAS 3987

8 BUF 3582

9 MIA 3314

10 DCA 3270

Problem 6 in the Chapter 15 Exercises (in Section 15.8 of the book).

The FAA code of the airport, full airport name, and the mean arrival delay figure in 2013 for each of the ten destinations that were worst in mean arrival delay are:

> sqldf("

+ SELECT

+ dest,

+ airports.name AS dest\_name,

+ AVG(arr\_delay) AS DelayMean

+ FROM flights

+ JOIN airports ON flights.dest = airports.faa

+ WHERE year = 2013

+ GROUP BY dest

+ ORDER BY DelayMean DESC

+ LIMIT 0, 10

+ ")

dest dest\_name DelayMean

1 CAE Columbia Metropolitan 41.76415

2 TUL Tulsa Intl 33.65986

3 OKC Will Rogers World 30.61905

4 JAC Jackson Hole Airport 28.09524

5 TYS Mc Ghee Tyson 24.06920

6 MSN Dane Co Rgnl Truax Fld 20.19604

7 RIC Richmond Intl 20.11125

8 CAK Akron Canton Regional Airport 19.69834

9 DSM Des Moines Intl 19.00574

10 GRR Gerald R Ford Intl 18.18956

**The airport with the highest mean arrival delay figure in 2013 is Columbia!!!**

Problem 7 in the Chapter 15 Exercises (in Section 15.8 of the book).

**111,279** flights in the 'flights' data table flew into JFK or flew out of JFK in 2013.

> sqldf("SELECT

+ COUNT(\*) AS numFlights

+ FROM flights

+ WHERE year = 2013 AND (origin = 'JFK' or dest = 'JFK')

+ LIMIT 0, 10

+ ")

numFlights

1 111279

Problem 8 in the Chapter 15 Exercises (in Section 15.8 of the book).

The airline (carrier) and flight number for all flights \*from\* JFK \*to\* LAX on September 26, 2013 are:

> sqldf("SELECT

+ carrier, flight

+ FROM flights

+ WHERE year = 2013 AND month = 9 and day = 26 AND (origin = 'JFK' AND dest = 'LAX')

+ ")

carrier flight

1 UA 797

2 B6 23

3 VX 399

4 DL 763

5 AA 33

6 UA 443

7 DL 120

8 VX 407

9 B6 223

10 AA 19

11 AA 1

12 UA 703

13 DL 863

14 B6 323

15 VX 411

16 AA 3

17 AA 117

18 UA 841

19 DL 1163

20 AA 181

21 B6 423

22 VX 413

23 DL 95

24 AA 133

25 UA 535

26 DL 17

27 VX 415

28 UA 771

29 B6 523

30 DL 2363

31 AA 21

32 AA 185

Problem 2 in the Chapter 16 Exercises (in Section 16.6 of the book).

Using a primary key for retroID may not be good, since some players may not appear in the other databases, and a primary key variable cannot have a NULL value. Using a unique key may be better, since unique key variables can have NULL values. Maybe a foreign key designation for retroID could be okay if retroID is a primary key in *another* database.