

Homework 10

Brady Lamson

2022-05-02

```
# Open a connection to the airlines database:
db_conn <- dbConnect_scidb("airlines")
```

First let's see what is even in flights for future reference.

```
dbGetQuery(conn = db_conn,
            statement = "DESCRIBE flights;"
)
```

##	Field	Type	Null	Key	Default	Extra
## 1	year	smallint(4)	YES	MUL	<NA>	
## 2	month	smallint(2)	YES		<NA>	
## 3	day	smallint(2)	YES		<NA>	
## 4	dep_time	smallint(4)	YES		<NA>	
## 5	sched_dep_time	smallint(4)	YES		<NA>	
## 6	dep_delay	smallint(4)	YES		<NA>	
## 7	arr_time	smallint(4)	YES		<NA>	
## 8	sched_arr_time	smallint(4)	YES		<NA>	
## 9	arr_delay	smallint(4)	YES		<NA>	
## 10	carrier	varchar(2)	NO	MUL		
## 11	tailnum	varchar(6)	YES	MUL	<NA>	
## 12	flight	smallint(4)	YES		<NA>	
## 13	origin	varchar(3)	NO	MUL		
## 14	dest	varchar(3)	NO	MUL		
## 15	air_time	smallint(4)	YES		<NA>	
## 16	distance	smallint(4)	YES		<NA>	
## 17	cancelled	tinyint(1)	YES		<NA>	
## 18	diverted	tinyint(1)	YES		<NA>	
## 19	hour	smallint(2)	YES		<NA>	
## 20	minute	smallint(2)	YES		<NA>	
## 21	time_hour	datetime	YES		<NA>	

Problem 2

```
dbGetQuery(conn = db_conn,  
           statement = "SELECT DISTINCT year  
                        FROM flights;"  
)
```

```
##   year  
## 1 2010  
## 2 2011  
## 3 2012  
## 4 2013  
## 5 2014  
## 6 2015  
## 7 2016  
## 8 2017
```

Exercise 3

```
dbGetQuery(conn = db_conn,  
           statement = "SELECT COUNT(*) AS number_of_flights  
                        FROM flights  
                        WHERE dest = 'DFW' AND year = 2010 AND month = 5 AND day = 14;"  
)
```

```
##   number_of_flights  
## 1                  754
```

Problem 5

```
dbGetQuery(conn = db_conn,  
           statement = "SELECT dest, COUNT(*) AS frequency  
                        FROM flights  
                        WHERE origin = 'ORD' AND year = 2010  
                        GROUP BY dest  
                        ORDER BY frequency DESC  
                        LIMIT 5;"  
)
```

```
##   dest frequency  
## 1  LGA        9787
```

```
## 2 DCA      7988
## 3 DFW      7468
## 4 LAX      7273
## 5 MSP      7062
```

Problem 6

```
dbGetQuery(conn = db_conn,
            statement = "SELECT dest, AVG(arr_delay) AS avg_arr_delay
                        FROM flights
                        WHERE year = 2010
                        GROUP BY dest
                        ORDER BY avg_arr_delay DESC
                        LIMIT 5;")
)
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 1 imported as
## numeric
```

```
##   dest avg_arr_delay
## 1 GUM      43.9000
## 2 ABR      19.4000
## 3 MOD      19.0308
## 4 OTH      18.5420
## 5 RDD      17.1446
```

Problem 8

```
dbGetQuery(conn = db_conn,
            statement = "SELECT carrier, flight
                        FROM flights
                        WHERE year = 1990 AND day = 26 AND month = 9 AND origin = 'LAX' AND dest = 'JFK'")
)
```

```
## [1] carrier flight
## <0 rows> (or 0-length row.names)
```

This returns 0 flights. This is both surprising and unsurprising. Unsurprising because our previous `SELECT DISTINCT year` query returned only flights from 2010 to 2017, but this is surprising because I don't know why this is a problem being asked in the first place.