# Homework 10

Tobias Boggess

2022-05-01

## Chapter 15 Problems

**Problem 2: Identify what years of data are available in the flights table of the airlines database.**

Code:

```
con <- dbConnect_scidb("airlines")
# con <- dbConnect(drv = MySQL(),
#                  dbname = "airlines",
#                  host = "mdsr.cdc7tgkkqd0n.us-east-1.rds.amazonaws.com",
#                  user = "mdsr_public",
#                  password = "ImhsmflMDSwR")

dbGetQuery(conn = con,
           statement = "SELECT year FROM flights
           GROUP BY year
           ORDER BY year")
```

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

The years available in the flights data are from 2010 to 2017.

**Problem 3: Use the dbConnect_scidb function to connect to the airlines database to answer the following problem. How many domestic flights flew into Dallas-Fort Worth (DFW) on May 14, 2010?**

Code:

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

```
##   count
## 1   754
```

There are 754 flights that flew into Dallas Fort Worth on May 14, 2010.

**Problem 5: Use the dbConnect_scidb function to connect to the airlines database to answer the following problem. Of all the destinations from Chicago O'Hare (ORD), which were the most common in 2010?**

Code:

```
dbGetQuery(conn = con,
           statement = "SELECT dest, origin, COUNT(*) AS count
           FROM flights
           WHERE year = 2010 AND origin = 'ORD'
           GROUP BY dest
           ORDER BY count DESC
           LIMIT 0,5;")
```

```
##   dest origin count
## 1  LGA    ORD  9787
## 2  DCA    ORD  7988
## 3  DFW    ORD  7468
## 4  LAX    ORD  7273
## 5  MSP    ORD  7062
```

The most popular destination from Chicago O'Hare airport was LGA with 9787 flights.

**Problem 6: Use the dbConnect_scidb function to connect to the airlines database to answer the following problem. Which airport had the highest average arrival delay time in 2010?**

Code:

```
dbGetQuery(conn = con,
           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 0,5;")
```

```
## 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
```

GUM airport had the highest average arrival delay time in 2010.

**Problem 8: Use the dbConnect_scidb function to connect to the airlines database to answer the following problem. List the airline and flight number for all flights between LAX and JFK on September 26th, 1990.**

Code:

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

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

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

The above doesn't return any flights because the flights dataset is only available between 2010 and 2017.