Relational data and database exercises - solution

Load packages and data

```r
library(tidyverse)
library(nycflights13)
```

Exercise 1: Average arrival delay for each carrier

What is the average arrival delay on flights for each carrier?

Show the results together with both carrier code and carrier name

(*Hint:* Use an inner join of `flights` and `airlines` and appropriate `group_by()` and `summarise()`.)

```r
flights |>
  inner_join(airlines, by = "carrier") |>
  group_by(carrier, name) |>
  summarise(`mean delay` = mean(arr_delay, na.rm = TRUE), .groups = "drop") |>
  collect() # Only needed if data is in a database outside R
```

```
# A tibble: 16 x 3
   carrier name                        `mean delay`
   <chr>   <chr>                              <dbl>
 1 9E      Endeavor Air Inc.                  7.38
 2 AA      American Airlines Inc.             0.364
 3 AS      Alaska Airlines Inc.              -9.93
 4 B6      JetBlue Airways                    9.46
 5 DL      Delta Air Lines Inc.               1.64
 6 EV      ExpressJet Airlines Inc.          15.8
 7 F9      Frontier Airlines Inc.            21.9
 8 FL      AirTran Airways Corporation       20.1
 9 HA      Hawaiian Airlines Inc.            -6.92
10 MQ      Envoy Air                         10.8
11 OO      SkyWest Airlines Inc.             11.9
12 UA      United Air Lines Inc.              3.56
13 US      US Airways Inc.                    2.13
14 VX      Virgin America                     1.76
15 WN      Southwest Airlines Co.             9.65
16 YV      Mesa Airlines Inc.                15.6
```

Exercise 2: Flights without records in `planes`

For each carrier, what's the number of flights that don't have matching records in `planes`?

Show results together with carrier code and name as before

Hint: We need three different tables and two different joins (have a look at `anti_join()` for the first part).

```r
flights |>
  anti_join(planes, by = "tailnum") |>
  inner_join(airlines, by = "carrier") |>
  count(carrier, name) |>
  collect() # Only needed if data is in a database outside R
```

```
# A tibble: 10 x 3
   carrier name                            n
   <chr>   <chr>                       <int>
 1 9E      Endeavor Air Inc.            1044
 2 AA      American Airlines Inc.      22558
 3 B6      JetBlue Airways               830
 4 DL      Delta Air Lines Inc.          110
 5 F9      Frontier Airlines Inc.         50
 6 FL      AirTran Airways Corporation   187
 7 MQ      Envoy Air                   25397
 8 UA      United Air Lines Inc.        1693
 9 US      US Airways Inc.               699
10 WN      Southwest Airlines Co.         38
```

Exercise 3: Number of flights by carrier and destination

What is the number of flights for each carrier and destination?

Show the name of both the carrier and airport, not the id codes, and sorted by airline and airport for clarity.

Hint: We need to join three different tables and notice that two tables have a variable `name` with different meanings so they will appear with two different suffixes (you can rename them if you like).

```
flights |>
  inner_join(airlines, by = "carrier") |>
  inner_join(airports, by = c("dest" = "faa")) |>
  rename(airline = name.x,
         airport = name.y) |>
  count(airline, airport) |>
  arrange(airline, airport) |>
  collect() # Only needed if data is in a database outside R
```

```
# A tibble: 304 x 3
   airline                     airport                                n
   <chr>                       <chr>                              <int>
 1 AirTran Airways Corporation Akron Canton Regional Airport        864
 2 AirTran Airways Corporation General Mitchell Intl                 59
 3 AirTran Airways Corporation Hartsfield Jackson Atlanta Intl     2337
 4 Alaska Airlines Inc.        Seattle Tacoma Intl                  714
 5 American Airlines Inc.      Austin Bergstrom Intl                365
 6 American Airlines Inc.      Chicago Ohare Intl                  6059
 7 American Airlines Inc.      Dallas Fort Worth Intl              7257
 8 American Airlines Inc.      Eagle Co Rgnl                        103
 9 American Airlines Inc.      Fort Lauderdale Hollywood Intl       182
10 American Airlines Inc.      General Edward Lawrence Logan Intl  1455
# i 294 more rows
```