NYC Flights 2013 Analysis

```
library(tidyverse)
```

```
flights <- read_csv("flights.csv")
airlines <- read_csv("airlines.csv")
airports <- read_csv("airports.csv")
```

```
Rows: 336776 Columns: 19

── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr   (4): carrier, tailnum, origin, dest
dbl  (14): year, month, day, dep_time, sched_dep_time, dep_delay, arr_time,
dttm  (1): time_hour


ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this mess

Rows: 16 Columns: 2

── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (2): carrier, name


ℹ Use `spec()` to retrieve the full column specification for this data.
```

```
glimpse(flights)
glimpse(airlines)
glimpse(airports)
```

```
Rows: 336,776
Columns: 19
$ year           <dbl> 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013,
$ month          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
$ day            <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
$ dep_time       <dbl> 517, 533, 542, 544, 554, 554, 555, 557, 557, 558, 558
$ sched_dep_time <dbl> 515, 529, 540, 545, 600, 558, 600, 600, 600, 600, 600
$ dep_delay      <dbl> 2, 4, 2, -1, -6, -4, -5, -3, -3, -2, -2, -2, -2, -2,
$ arr_time       <dbl> 830, 850, 923, 1004, 812, 740, 913, 709, 838, 753, 84
$ sched_arr_time <dbl> 819, 830, 850, 1022, 837, 728, 854, 723, 846, 745, 85
$ arr_delay      <dbl> 11, 20, 33, -18, -25, 12, 19, -14, -8, 8, -2, -3, 7,
```

```
$ carrier        <chr> "UA", "UA", "AA", "B6", "DL", "UA", "B6", "EV", "B6",
$ flight         <dbl> 1545, 1714, 1141, 725, 461, 1696, 507, 5708, 79, 301,
$ tailnum        <chr> "N14228", "N24211", "N619AA", "N804JB", "N668DN", "N3
$ origin         <chr> "EWR", "LGA", "JFK", "JFK", "LGA", "EWR", "EWR", "LGA
$ dest           <chr> "IAH", "IAH", "MIA", "BQN", "ATL", "ORD", "FLL", "IAD
$ air_time       <dbl> 227, 227, 160, 183, 116, 150, 158, 53, 140, 138, 149,
$ distance       <dbl> 1400, 1416, 1089, 1576, 762, 719, 1065, 229, 944, 733
$ hour           <dbl> 5, 5, 5, 5, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6
```

Q1: Which carrier had the most flights in July 2013

Ans: United Air Lines Inc. (5,066 fights)

```
flights %>%
  filter(month == 7, year == 2013) %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name) %>%
  summarise(flights = n()) %>%
  arrange(desc(flights)) %>%
  rename(carrier_name = name)
```

A tibble: 15 × 2

carrier_name	flights
<chr>	<int>
United Air Lines Inc.	5066
JetBlue Airways	4984
ExpressJet Airlines Inc.	4641
Delta Air Lines Inc.	4251
American Airlines Inc.	2882
Envoy Air	2261
US Airways Inc.	1786
Endeavor Air Inc.	1494
Southwest Airlines Co.	1076
Virgin America	489
AirTran Airways Corporation	263
Mesa Airlines Inc.	81
Alaska Airlines Inc.	62
Frontier Airlines Inc.	58
Hawaiian Airlines Inc.	31

Q2: Which month had the fewest flights in 2013?

Ans: February 2013 (24,951 fights)

```
flights %>%
  group_by(month) %>%
  summarise(flights = n()) %>%
  arrange(flights)
```

A tibble: 12 × 2

month	flights
<dbl>	<int>
2	24951
1	27004
11	27268
9	27574
12	28135
6	28243
4	28330
5	28796
3	28834
10	28889
8	29327
7	29425

Q3: Which carrier had the most departure delays in 2013?

Ans: ExpressJet Airlines Inc. Total 1,164,581 minutes in 2013

```
flights %>%
  filter(dep_delay > 0) %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name) %>%
  rename(carrier_name = name) %>%
  summarise(sum_dep_delay = sum(dep_delay),
            n_dep_delay = n(),
            avg_dep_delay = round(mean(dep_delay), 0)) %>%
  arrange(desc(sum_dep_delay))
```

A tibble: 16 × 4

carrier_name	sum_dep_delay	n_dep_delay	avg_dep_delay
<chr>	<dbl>	<int>	<dbl>
ExpressJet Airlines Inc.	1164581	23139	50
JetBlue Airways	853387	21445	40
United Air Lines Inc.	815818	27261	30
Delta Air Lines Inc.	570017	15241	37
American Airlines Inc.	377714	10162	37
Envoy Air	360715	8031	45
Endeavor Air Inc.	345522	7063	49
Southwest Airlines Co.	228595	6558	35
US Airways Inc.	157817	4775	33
Virgin America	76662	2225	34
AirTran Airways Corporation	67526	1654	41
Frontier Airlines Inc.	15392	341	45
Mesa Airlines Inc.	12338	233	53
Alaska Airlines Inc.	7083	226	31
Hawaiian Airlines Inc.	3094	69	45
SkyWest Airlines Inc.	522	9	58

Q4: The 5 longest flights by distance in 2013?

Ans.

1. John F Kennedy Intl to Honolulu Intl (4,983 miles)
2. Newark Liberty Intl to Honolulu Intl (4,963 miles)
3. Newark Liberty Intl to Ted Stevens Anchorage Intl (3,370 miles)
4. John F Kennedy Intl to San Francisco Intl (2,586 miles)
5. John F Kennedy Intl to Metropolitan Oakland Intl (2,586 miles)

```
flights %>%
        left_join(airports, by = c("origin" = "faa")) %>%
        rename(origin_airport = name) %>%
        left_join(airports, by = c("dest" = "faa")) %>%
        rename(dest_airport = name) %>%
        select(origin_airport, dest_airport, distance) %>%
        distinct(origin_airport, dest_airport, distance) %>%
        arrange(desc(distance)) %>%
        head(5)
```

A tibble: 5 × 3

origin_airport	dest_airport	distance
<chr>	<chr>	<dbl>
John F Kennedy Intl	Honolulu Intl	4983
Newark Liberty Intl	Honolulu Intl	4963
Newark Liberty Intl	Ted Stevens Anchorage Intl	3370
John F Kennedy Intl	San Francisco Intl	2586
John F Kennedy Intl	Metropolitan Oakland Intl	2576

Q5: The 5 shortest flights by distance in 2013?

Ans.

1. Newark Liberty Intl to La Guardia (17 miles)
2. Newark Liberty Intl to Philadephia Intl (80 miles)
3. John F Kennedy Intl to Philadephia Intl (94 miles)
4. La Guardia to Philadephia Intl (96 miles)
5. Newark Liberty Intl to Bradley Intl (116 miles)

```
flights %>%
        left_join(airports, by = c("origin" = "faa")) %>%
        rename(origin_airport = name) %>%
        left_join(airports, by = c("dest" = "faa")) %>%
        rename(dest_airport = name) %>%
        select(origin_airport, dest_airport, distance) %>%
        distinct(origin_airport, dest_airport, distance) %>%
        arrange(distance) %>%
        head(5)
```

A tibble: 5 × 3

origin_airport	dest_airport	distance
<chr>	<chr>	<dbl>
Newark Liberty Intl	La Guardia	17
Newark Liberty Intl	Philadelphia Intl	80
John F Kennedy Intl	Philadelphia Intl	94
La Guardia	Philadelphia Intl	96
Newark Liberty Intl	Bradley Intl	116