NYC Flights 2013 Analysis

```
library(tidyverse)
library(glue)
library(dplyr)
```

```
Warning message in system("timedatectl", intern = TRUE):
“running command 'timedatectl' had status 1”
Warning message:
“Failed to locate timezone database”
── Attaching packages ─────────────────────────────────────── tidyverse 1.3

✓ ggplot2 3.3.5     ✓ purrr   0.3.4
✓ tibble  3.1.5     ✓ dplyr   1.0.7
✓ tidyr   1.1.4     ✓ stringr 1.4.0
✓ readr   2.0.2     ✓ forcats 0.5.1

── Conflicts ────────────────────────────────────────── tidyverse_conflicts
✖ dplyr::filter()  masks stats::filter()
✖ purrr::flatten() masks jsonlite::flatten()
✖ dplyr::lag()     masks stats::lag()


Attaching package: ‘glue’
```

```
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 mes

Rows: 16 Columns: 2

── Column specification ──────────────────────────────────────────────────
Delimiter: ","
```

```
chr (2): carrier, name
```

```
glimpse(flights)
```

```
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, 1
$ day            <dbl> 1, 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, 55
$ sched_dep_time <dbl> 515, 529, 540, 545, 600, 558, 600, 600, 600, 600, 60
$ 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, 8
$ sched_arr_time <dbl> 819, 830, 850, 1022, 837, 728, 854, 723, 846, 745, 8
$ 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", "N
$ origin         <chr> "EWR", "LGA", "JFK", "JFK", "LGA", "EWR", "EWR", "LG
$ dest           <chr> "IAH", "IAH", "MIA", "BQN", "ATL", "ORD", "FLL", "IA
$ 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, 73
$ hour           <dbl> 5, 5, 5, 5, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6
$ minute         <dbl> 15, 29, 40, 45, 0, 58, 0, 0, 0, 0, 0, 0, 0, 0, 0, 59
```

```
clean_flights <- drop_na(flights)
clean_flights %>%
filter(is.na(clean_flights))
```

A tibble: 0 × 19

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl

```
glimpse(airlines)
```

```
Rows: 16
Columns: 2
$ carrier <chr> "9E", "AA", "AS", "B6", "DL", "EV", "F9", "FL", "HA", "MQ",
$ name    <chr> "Endeavor Air Inc.", "American Airlines Inc.", "Alaska Airli
```

```
airlines %>%
filter(is.na(airlines))
```

A spec_tbl_df:
0 × 2

carrier	name
<chr>	<chr>

```
glimpse(airports)
```

```
Rows: 1,458
Columns: 8
$ faa   <chr> "04G", "06A", "06C", "06N", "09J", "0A9", "0G6", "0G7", "0P2",
$ name  <chr> "Lansdowne Airport", "Moton Field Municipal Airport", "Schaumb
$ lat   <dbl> 41.13047, 32.46057, 41.98934, 41.43191, 31.07447, 36.37122, 41
$ lon   <dbl> -80.61958, -85.68003, -88.10124, -74.39156, -81.42778, -82.173
$ alt   <dbl> 1044, 264, 801, 523, 11, 1593, 730, 492, 1000, 108, 409, 875,
$ tz    <dbl> -5, -6, -6, -5, -5, -5, -5, -5, -5, -8, -5, -6, -5, -5, -5, -5
$ dst   <chr> "A", "A", "A", "A", "A", "A", "A", "A", "U", "A", "A", "U", "A
$ tzone <chr> "America/New_York", "America/Chicago", "America/Chicago", "Ame
```

Q1: What are the top 5 airlines with the combined highest number of delayed arrival and departure minutes?

```
clean_flights %>%
left_join(airlines, "carrier") %>%
filter(arr_delay > 0, dep_delay > 0) %>%
mutate(sum_arrdep_delay = arr_delay + dep_delay) %>%
select(airline_name = name, sum_arrdep_delay) %>%
count(airline_name) %>%
arrange(desc(n)) %>%
rename(sum_min_arrdep_delay = n) %>%
head(5)
```

A tibble: 5 × 2

airline_name	sum_min_arrdep_delay
<chr>	<int>
ExpressJet Airlines Inc.	19183
United Air Lines Inc.	16606
JetBlue Airways	16436
Delta Air Lines Inc.	10126
Envoy Air	6944

Q2: What were the top 5 destination on Christmas?

```
clean_flights %>%
left_join(airports, by = c("dest" = "faa")) %>%
filter(day == 25, month == 12) %>%
count(destination = name) %>%
arrange(desc(n)) %>%
head(5)
```

A tibble: 5 × 2

destination	n
<chr>	<int>
Orlando Intl	41
Fort Lauderdale Hollywood Intl	39
Hartsfield Jackson Atlanta Intl	37
Los Angeles Intl	36
Charlotte Douglas Intl	32

Q3: Which top 5 airline had the most flights in 2013?

```
clean_flights %>%
left_join(airlines, "carrier") %>%
filter(year == 2013) %>%
group_by(airline_name = name) %>%
summarise(sum_num_flight = sum(flight)) %>%
arrange(desc(sum_num_flight)) %>%
head(5)
```

A tibble: 5 × 2

airline_name	sum_num_flight
<chr>	<dbl>
ExpressJet Airlines Inc.	236289047
Envoy Air	96562720
Delta Air Lines Inc.	65485862
Endeavor Air Inc.	61608821
United Air Lines Inc.	55574781

Q4: On average, Which airport is the earliest to fly to?

```
clean_flights %>%
left_join(airports, c("dest" = "faa")) %>%
group_by(airport_name = name) %>%
summarise(avg_air_time = mean(air_time)) %>%
arrange(avg_air_time) %>%
head(1)
```

A tibble: 1 × 2

airport_name	avg_air_time
<chr>	<dbl>
Bradley Intl	25.46602

Q5: Top 5 furthest airports

```r
clean_flights %>%
left_join(airports, c("dest" = "faa")) %>%
distinct(airport_name = name, distance) %>%
arrange(desc(distance)) %>%
head(5)
```

A tibble: 5 × 2

distance	airport_name
<dbl>	<chr>
4983	Honolulu Intl
4963	Honolulu Intl
3370	Ted Stevens Anchorage Intl
2586	San Francisco Intl
2576	Metropolitan Oakland Intl