NYC Flights 2013 Analysis

```
#load library
library(tidyverse)
```

```
Warning message in system("timedatectl", intern = TRUE):
“running command 'timedatectl' had status 1”
Warning message:
“Failed to locate timezone database”
── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ─

✓ 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()
```

```
flights <- read.csv("flights.csv")
airlines <- read.csv("airlines.csv")
```

```
glimpse(flights)
```

```
Rows: 336,776
Columns: 19
$ year           <int> 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2
$ month          <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
$ day            <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
$ dep_time       <int> 517, 533, 542, 544, 554, 554, 555, 557, 557, 558, 558,
$ sched_dep_time <int> 515, 529, 540, 545, 600, 558, 600, 600, 600, 600, 600,
$ dep_delay      <int> 2, 4, 2, -1, -6, -4, -5, -3, -3, -2, -2, -2, -2, -2, -1
$ arr_time       <int> 830, 850, 923, 1004, 812, 740, 913, 709, 838, 753, 849,
```

```
$ sched_arr_time <int> 819, 830, 850, 1022, 837, 728, 854, 723, 846, 745, 851,
$ arr_delay      <int> 11, 20, 33, -18, -25, 12, 19, -14, -8, 8, -2, -3, 7, -1
$ carrier        <chr> "UA", "UA", "AA", "B6", "DL", "UA", "B6", "EV", "B6", "
$ flight         <int> 1545, 1714, 1141, 725, 461, 1696, 507, 5708, 79, 301, 4
$ tailnum        <chr> "N14228", "N24211", "N619AA", "N804JB", "N668DN", "N394
$ origin         <chr> "EWR", "LGA", "JFK", "JFK", "LGA", "EWR", "EWR", "LGA",
$ dest           <chr> "IAH", "IAH", "MIA", "BQN", "ATL", "ORD", "FLL", "IAD",
$ air_time       <int> 227, 227, 160, 183, 116, 150, 158, 53, 140, 138, 149, 1
$ distance       <int> 1400, 1416, 1089, 1576, 762, 719, 1065, 229, 944, 733,
$ hour           <int> 5, 5, 5, 5, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 5, 6, 6, 6
```

```
glimpse(airlines)
```

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

Q1 : How many flights destination at LAX on 25 December 2013?

```
flights %>%
    filter(dest == "LAX", month == 12, day == 25) %>%
    count(dest)
```

A data.frame: 1 × 2

dest	n
<chr>	<int>
LAX	36

Q2 : How many flights in 2013 group by

carrier?

```
flights %>%
    group_by(carrier) %>%
    summarise(count_flight = n()) %>%
    arrange(desc(count_flight)) %>%
    left_join(airlines ,by = "carrier")
```

A tibble: 16 × 3

carrier	count_flight	name
<chr>	<int>	<chr>
UA	58665	United Air Lines Inc.
B6	54635	JetBlue Airways
EV	54173	ExpressJet Airlines Inc.
DL	48110	Delta Air Lines Inc.
AA	32729	American Airlines Inc.
MQ	26397	Envoy Air
US	20536	US Airways Inc.
9E	18460	Endeavor Air Inc.
WN	12275	Southwest Airlines Co.
VX	5162	Virgin America
FL	3260	AirTran Airways Corporation
AS	714	Alaska Airlines Inc.
F9	685	Frontier Airlines Inc.
YV	601	Mesa Airlines Inc.
HA	342	Hawaiian Airlines Inc.
OO	32	SkyWest Airlines Inc.

Q:3 Top 10 airlines had the most flights in June 2013

```
flights %>%

    filter(month == 6) %>%
    count(carrier) %>%
    arrange(desc(n)) %>%
```

```
    left_join(airlines, by = "carrier") %>%
    head(10)
```

A data.frame: 10 × 3

	carrier	n	name
	<chr>	<int>	<chr>
1	UA	4975	United Air Lines Inc.
2	B6	4622	JetBlue Airways
3	EV	4456	ExpressJet Airlines Inc.
4	DL	4126	Delta Air Lines Inc.
5	AA	2757	American Airlines Inc.
6	MQ	2178	Envoy Air
7	US	1736	US Airways Inc.
8	9E	1437	Endeavor Air Inc.
9	WN	1028	Southwest Airlines Co.
10	VX	480	Virgin America

Q4 : Top 3 flights have the longest distance

```
flights %>%
    distinct(carrier,flight, origin , dest, distance)%>%
    arrange(desc(distance))%>%
    head(3)
```

A data.frame: 3 × 5

	carrier	flight	origin	dest	distance
	<chr>	<int>	<chr>	<chr>	<int>
1	HA	51	JFK	HNL	4983
2	UA	15	EWR	HNL	4963
3	UA	887	EWR	ANC	3370

Q5 : What month has the most flights?

```
flights %>%
    group_by(month) %>%
    summarise(count_flight = n()) %>%
    arrange(desc(count_flight)) %>%
    head(1)
```

A tibble: 1 × 2

month	count_flight
<int>	<int>
7	29425