

NYC Flights 2013 Analysis

```
library(dplyr)
library(readr)
```

```
flights <- read_csv("flights.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
dtm     (1): time_hour

```

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

```
glimpse(flights)
```

```

Rows: 336,776
Columns: 19
$ year      <dbl> 2013, 2013, 2013, 2013, 2013, 2013, 2013, 2013, 20
$ month     <dbl> 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,
$ dep_time  <dbl> 517, 533, 542, 544, 554, 554, 555, 557, 557, 558,
$ sched_dep_time <dbl> 515, 529, 540, 545, 600, 558, 600, 600, 600, 600,
$ dep_delay <dbl> 2, 4, 2, -1, -6, -4, -5, -3, -3, -2, -2, -2, -2, -
$ arr_time  <dbl> 830, 850, 923, 1004, 812, 740, 913, 709, 838, 753,
$ sched_arr_time <dbl> 819, 830, 850, 1022, 837, 728, 854, 723, 846, 745,
$ arr_delay <dbl> 11, 20, 33, -18, -25, 12, 19, -14, -8, 8, -2, -3,
$ carrier   <chr> "UA", "UA", "AA", "B6", "DL", "UA", "B6", "EV", "E
$ flight    <dbl> 1545, 1714, 1141, 725, 461, 1696, 507, 5708, 79, 3
$ tailnum   <chr> "N14228", "N24211", "N619AA", "N804JB", "N668DN",
$ origin    <chr> "EWR", "LGA", "JFK", "JFK", "LGA", "EWR", "EWR",
$ dest      <chr> "IAH", "IAH", "MIA", "BQN", "ATL", "ORD", "FLL",
$ air_time  <dbl> 227, 227, 160, 183, 116, 150, 158, 53, 140, 138,

```

```
$ distance      <dbl> 1400, 1416, 1089, 1576, 762, 719, 1065, 229, 944,  
$ hour          <dbl> 5, 5, 5, 5, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 5, 6,
```

Q1 : What date had the most flight in NYC 2013?

```
flights %>%  
  count(month, day) %>%  
  arrange(desc(n)) %>%  
  rename(count_flight = n) %>%  
  head(5)
```

A tibble: 5 × 3

month	day	count_flight
<dbl>	<dbl>	<int>
11	27	1014
7	11	1006
7	8	1004
7	10	1004
12	2	1004

Q2 : Which carrier and what date had the most arrival delay?

```
flights %>%  
  select(day, month, year, carrier, arr_delay, origin, dest) %>%  
  arrange(desc(arr_delay)) %>%  
  head(5)
```

A tibble: 5 × 7

day	month	year	carrier	arr_delay	origin	dest
<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<chr>	<chr>
9	1	2013	HA	1272	JFK	HNL
15	6	2013	MQ	1127	JFK	CMH
10	1	2013	MQ	1109	EWB	ORD
20	9	2013	AA	1007	JFK	SFO
22	7	2013	MQ	989	JFK	CVG

Q3 : Which carrier had most flights in Nov 2013?

```
flights %>%  
  filter(month == 11, year == 2013) %>%  
  count(carrier) %>%  
  arrange(desc(n)) %>%  
  head(5)
```

A tibble: 5 × 2

carrier	n
<chr>	<int>
UA	4854
EV	4471
B6	4289
DL	3849
AA	2577

Q4 : Where origin and destination have the most distance ?

```
flights %>%  
  select(origin, dest, distance) %>%  
  arrange(desc(distance)) %>%  
  head(1)
```

A tibble: 1 × 3

origin	dest	distance
<chr>	<chr>	<dbl>
JFK	HNL	4983

Q5 : Which carrier had the most left from origin in EWR?

```
flights %>%  
  filter(origin == "EWR") %>%  
  count(carrier) %>%  
  arrange(desc(n)) %>%  
  head(5)
```

A tibble: 5 × 2

carrier	n
<chr>	<int>
UA	46087
EV	43939
B6	6557
WN	6188
US	4405