```
# shift + enter ==> RUN CODE
# ctrl + M ==> convert code cell to markdown cell
```

NYC Flights 2013 Analysis

```
library(dplyr)
```

```
Attaching package: ‘dplyr’


The following objects are masked from ‘package:stats’:

    filter, lag


The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union
```

```
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_
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 thi
```

```
head(flights)
```

A tibble: 6 × 19

year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	fli
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<(
2013	1	1	517	515	2	830	819	11	UA	15
2013	1	1	533	529	4	850	830	20	UA	17
2013	1	1	542	540	2	923	850	33	AA	1(
2013	1	1	544	545	-1	1004	1022	-18	B6	72
2013	1	1	554	600	-6	812	837	-25	DL	46
2013	1	1	554	558	-4	740	728	12	UA	16

1. How many flights are there in each month?

```
flights %>%
    count(month)
```

A spec_tbl_df: 12 × 2

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

2. How many flights are there in 2013?

```
flights %>%
    summarise(n())
```

A tibble: 1 × 1

n()
<int>
336776

3. Top 5 destination

```
flights %>%
    count(dest) %>%
    arrange(desc(n)) %>%
    head(5)
```

A tibble: 5 × 2

dest	n
<chr>	<int>
ORD	17283
ATL	17215
LAX	16174
BOS	15508
MCO	14082

4. How many flights are there in each carrier on Christmas?

```
flights %>%
    select(carrier, month, day) %>%
    filter(month == 12, day == 25) %>%
    count(carrier) %>%
    arrange(desc(n))
```

A tibble: 14 × 2

carrier	n
<chr>	<int>
B6	159
UA	121
DL	105
AA	78
EV	75
MQ	58
US	36
9E	32
WN	30
VX	13
FL	7
AS	2
F9	2
HA	1

5. How many flights delay in each carrier in 2013?

```
delay_flight <- data.frame(flights %>%
    select(carrier, flight, dep_delay) %>%
    mutate(delay = factor(if_else(dep_delay < 0 , T, F))) %>%
    filter(delay == T) %>%
    count(carrier) %>%
    arrange(desc(n)) %>%
    rename( no_of_delay_flight = n))
```

```
total_flight <- data.frame(flights %>%
    count(carrier) %>%
    arrange(desc(n)) %>%
    rename( no_of_flight = n))
```

```
delay_flight %>%
    left_join(total_flight, by = "carrier") %>%
    select(1, 3, 2) %>%
    mutate(percentage_delay = (no_of_delay_flight/no_of_flight)*100) %>%
    arrange(desc(percentage_delay))
```

A data.frame: 16 × 4

carrier	no_of_flight	no_of_delay_flight	percentage_delay
<chr>	<int>	<int>	<dbl>
HA	342	259	75.73099
US	20536	14461	70.41780
AS	714	458	64.14566
OO	32	20	62.50000
AA	32729	20324	62.09783
DL	48110	29654	61.63791
MQ	26397	15825	59.94999
B6	54635	29952	54.82200
9E	18460	9718	52.64355
EV	54173	26558	49.02442
YV	601	294	48.91847
VX	5162	2495	48.33398
UA	58665	27321	46.57121
F9	685	297	43.35766
FL	3260	1401	42.97546
WN	12275	4538	36.96945