

# DATA Analysis with R Programming

**Author :** Kamolchanok T.

**Course :** Data Transformation with Datarockie

**Description :** use `dplyr` to analyze `nycflights13` dataframe.

```
library(dplyr)
```

```
# import nycflights13 csv files
airports <- read.csv("airports.csv")
airplanes <- read.csv("planes.csv")
airlines <- read.csv("airlines.csv")
weather <- read.csv("weather.csv")
flights <- read.csv("flights.csv")
```

## QUESTION 1

Top 5 Destination Airport

```
flights %>%
  left_join(airports, by = c("dest" = "faa")) %>%
  group_by(name) %>%
  count(dest, sort=TRUE) %>%
  select("Airport Name"="name",
         "Airport Code"= "dest",
         "No.of flight"="n") %>%
  head(5)
```

A grouped\_df: 5 × 3

Airport Name	Airport Code	No.of flight
<chr>	<chr>	<int>
Chicago Ohare Intl	ORD	17283
Hartsfield Jackson Atlanta Intl	ATL	17215
Los Angeles Intl	LAX	16174
General Edward Lawrence Logan Intl	BOS	15508
Orlando Intl	MCO	14082

## QUESTION 2

Top 5 Delay Airlines in 2013 (Minute)

```
delay <- flights %>%
  filter(!is.na(arr_delay) & arr_delay > 0) %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name) %>%
  summarise(total_min_delay = sum(arr_delay)) %>%
  arrange(desc(total_min_delay))

names(delay)[1] = "Airline Name"

head(delay, 5)
```

A tibble: 5 × 2

Airline Name	total_min_delay
<chr>	<int>
ExpressJet Airlines Inc.	1181808
JetBlue Airways	944574
United Air Lines Inc.	814458
Delta Air Lines Inc.	619485
Envoy Air	442604

## QUESTION 3

Which airlines use the oldest plane in operation?

```
pf <- airplanes %>% left_join(flights, by = "tailnum")

oldest <- pf %>%
  left_join(airlines, by = "carrier") %>%
  select(name,model,tailnum,year.x) %>%
  group_by(model) %>%
  arrange(year.x) %>%
  unique()

oldest[1,]
```

A grouped\_df: 1 × 4

name	model	tailnum	year.x
<chr>	<chr>	<chr>	<int>
American Airlines Inc.	DC-7BF	N381AA	1956

## QUESTION 4

What's the hottest temperature for each origin?

```
aw <- airports %>% select(origin = faa,name)

w <- weather %>%
  group_by(origin) %>%
  summarise(max_temp = max(temp, na.rm=TRUE))

weather1 <- aw %>% right_join(w, by = "origin")
colnames(weather1)[c(1,2,3)] = c("Airport Code","Airport Name","Max Temp")
weather1
```

A data.frame: 3 × 3

Airport Code	Airport Name	Max Temp
<chr>	<chr>	<dbl>
EWR	Newark Liberty Intl	100.04
JFK	John F Kennedy Intl	98.06
LGA	La Guardia	98.96

## QUESTION 5

What's the airplane operate the longest total distance ?

```
flights %>%  
  filter(!is.na(tailnum)) %>%  
  group_by(tailnum) %>%  
  summarise(total_distance = sum(distance)) %>%  
  arrange(desc(total_distance)) %>%  
  head(1)
```

A tibble: 1 × 2

tailnum	total_distance
<chr>	<int>
N328AA	939101