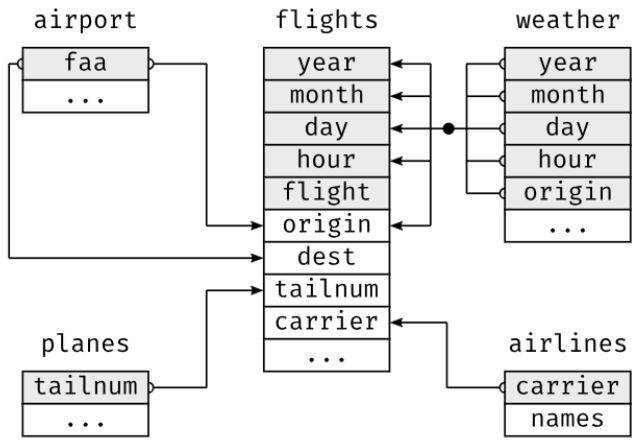
Data Analysis With R

Using **Airline on-time data for all flights departing NYC in 2013** dataset from “nycflights13” library

Database schema :



🔍 Loading library ⇒

```
library(nycflights13)
```

```
library(tidyverse)
data(”flights”)
```

Q.1 → Total flight in Y2013 (Only complete trip data <exclude “NA”>)

```
library(nycflights13)
library(tidyverse)
library(glue)
data("flights")

flights %>%
  filter(dep_time != "NA") %>%
  summarise(n=n()) -> tol
  glue("Total flight in Y2013 = {tol} flights")
```

```
>   glue("Total flight in Y2013 = {tol} flights")
Total flight in Y2013 = 328521 flights
```

Q.2 → Which carrier have most travel distance (only 1 way)

```
library(nycflights13)
library(tidyverse)
library(glue)
data("flights")

d_flight <- flights %>%
              select(carrier,distance) %>%
              group_by(carrier) %>%
              summarise(maxdis_mil=max(distance)) %>%
              mutate(maxdis_km = maxdis_mil*1.61) %>%
              arrange(desc(maxdis_km))
glue("Most travel distance carrier is {head(d_flight$carrier,1)} by {head(d_flight$maxdis_km,1)}
```

```
> glue("Most travel distance carrier is {head(d_flight$carrier,1)} by {h
ead(d_flight$maxdis_km,1)} km")
Most travel distance carrier is HA by 8022.63 km
```

Q.3 → Which carrier & tail number plane have most departure delayed time

```
library(nycflights13)
library(tidyverse)
library(glue)
data("flights")

d_flight_2 <- flights %>%
              select(carrier,tailnum,dep_delay) %>%
              mutate(dep_delay_hr = round(dep_delay/60,2)) %>%
              arrange(desc(dep_delay))
glue("Most delyed flight carrier is {head(d_flight_2$carrier,1)} & Plane number is {head(d_fligh
```

```
> glue("Most delyed flight carrier is {head(d_flight_2$carrier,1)} & Pla
ne number is {head(d_flight_2$tailnum,1)} by {head(d_flight_2$dep_delay_
hr,1)} hour")
Most delyed flight carrier is HA & Plane number is N384HA by 21.68 hour
```

Q.4 → Top 3 arrival destination (airport) in Y2013

```
library(nycflights13)
library(tidyverse)
library(glue)
data("flights")

d_flight_3 <- flights %>%
              select(dest) %>%
              count(dest, sort=TRUE) %>%
              group_by(dest)
glue("Top 3 delyed flight carrier :","
      No.1 carrier is {d_flight_3[1,1]} or O'Hare International Airport
      No.2 carrier is {d_flight_3[2,1]} or Atlanta International Airport
      No.3 carrier is {d_flight_3[3,1]} or Los Angeles International Airport ")
```

```
Top 3 delyed flight carrier :
No.1 carrier is ORD or O'Hare International Airport
No.2 carrier is ATL or Atlanta International Airport
No.3 carrier is LAX or Los Angeles International Airport
```

Q.5 → Which month in Y2013 have lowest flight (by count)

```
library(nycflights13)
library(tidyverse)
library(glue)
data("flights")

d_flight_4 <- flights %>%
              select(month,flight,dep_time) %>%
              filter(dep_time>0) %>%
              group_by(month) %>%
              summarise(flight_count=n())
glue("Lowest flight are in month : {d_flight_4[12,1]}, Total flight is {d_flight_4[12,2]} flight
```

```
> glue("Lowest flight are in month : {d_flight_4[12,1]}, Total flight is
{d_flight_4[12,2]} flight ")
Lowest flight are in month : 2, Total flight is 23690 flight
```