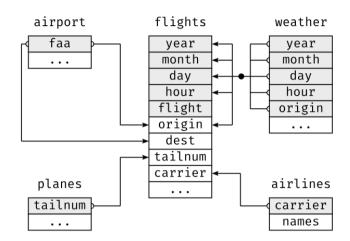
Data Analysis With R

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)

```
> 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
```

Data Analysis With R

1

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

```
> 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

```
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)

```
> 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
```

Data Analysis With R

2