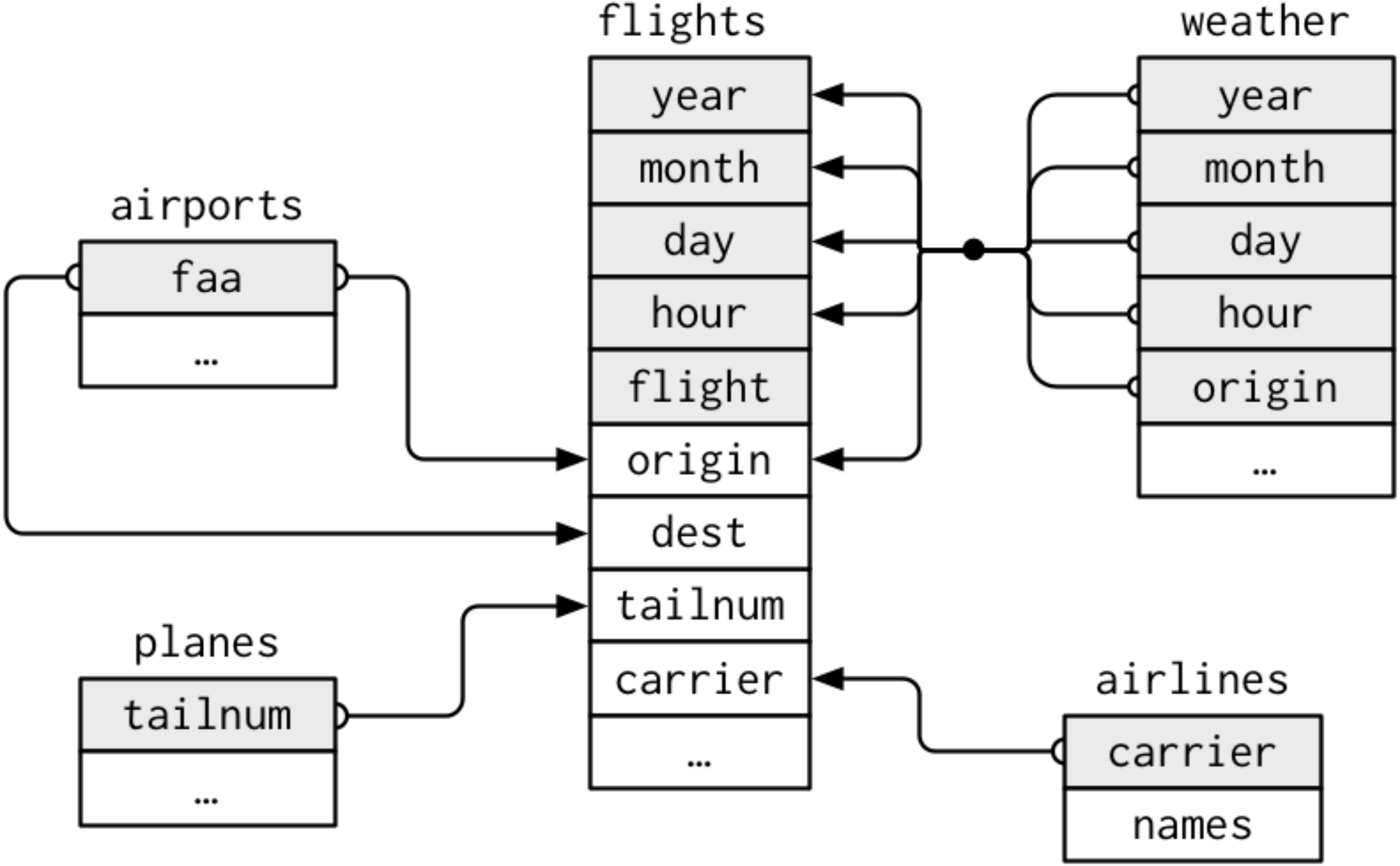
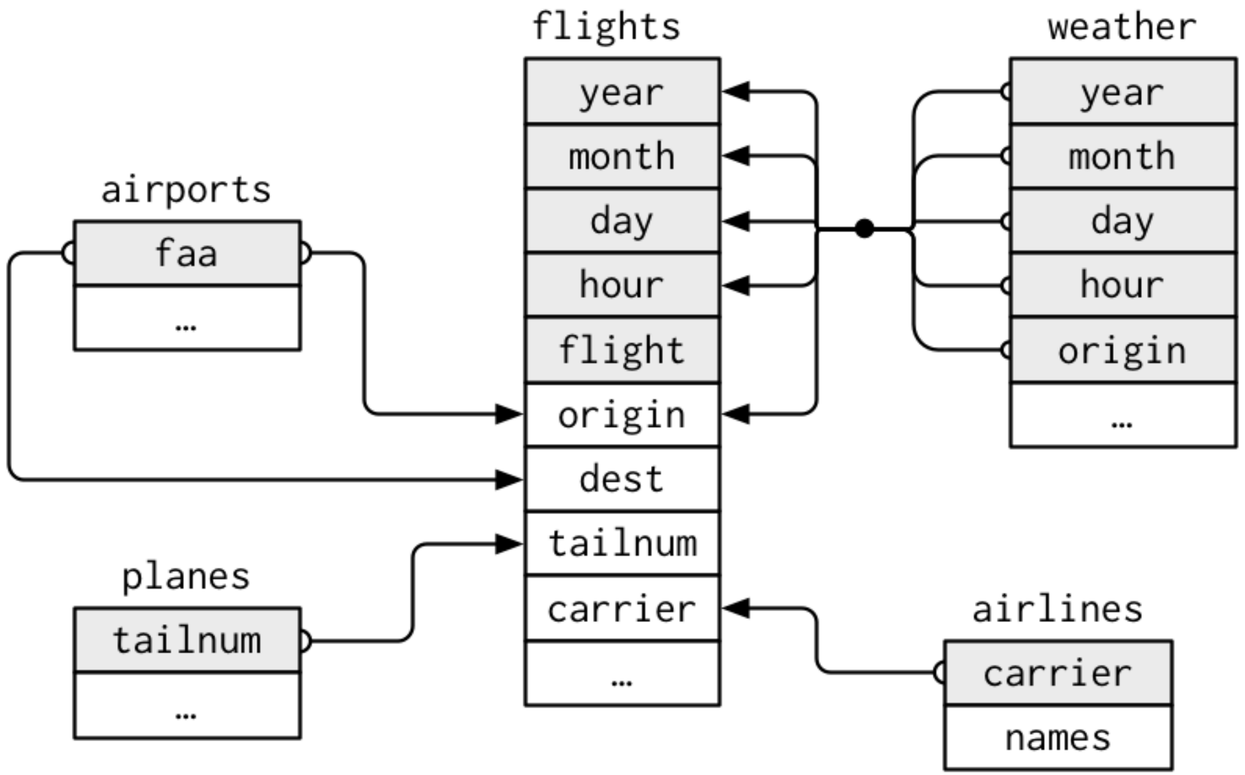
Tables linked by primary and foreign keys form a **relational database**


flights
year
month
day
hour
flight
origin
dest
tailnum
carrier
…
weather
year
month
day
hour
origin
…
airports
faa
…
planes
tailnum
…
airlines
carrier
names

A primary key and the corresponding foreign key in another table form a *relation*

Relations are often 1-to-many
(e.g. each flight has one plane, but each plane has many flights)

Tables linked by primary and foreign keys form a **relational database**



A primary key and the corresponding foreign key in another table form a *relation*

Relations are often 1-to-many
(e.g. each flight has one plane, but each plane has many flights)

Problem: Integrate data from multiple tables

Solution: Mutating joins

```
flights2 <- flights %>%
  select(year:day, hour, origin, dest, tailnum, carrier)
flights2
#> # A tibble: 336,776 x 8
#>    year month   day  hour origin dest  tailnum carrier
#>   <int> <int> <int> <dbl> <chr>  <chr> <chr>   <chr>
#> 1  2013     1     1     5 EWR    IAH   N14228  UA
#> 2  2013     1     1     5 LGA    IAH   N24211  UA
#> 3  2013     1     1     5 JFK    MIA   N619AA  AA
#> 4  2013     1     1     5 JFK    BQN   N804JB  B6
#> 5  2013     1     1     6 LGA    ATL   N668DN  DL
#> 6  2013     1     1     5 EWR    ORD   N39463  UA
#> # … with 3.368e+05 more rows
```

```
flights2 %>%
  select(-origin, -dest) %>%
  left_join(airlines, by = "carrier")
#> # A tibble: 336,776 x 7
#>    year month   day  hour tailnum carrier name
#>   <int> <int> <int> <dbl> <chr>   <chr>   <chr>
#> 1  2013     1     1     5 N14228  UA      United Air Lines Inc.
#> 2  2013     1     1     5 N24211  UA      United Air Lines Inc.
#> 3  2013     1     1     5 N619AA  AA      American Airlines Inc.
#> 4  2013     1     1     5 N804JB  B6      JetBlue Airways
#> 5  2013     1     1     6 N668DN  DL      Delta Air Lines Inc.
#> 6  2013     1     1     5 N39463  UA      United Air Lines Inc.
#> # … with 3.368e+05 more rows
```