Lab 11: Databases

MinJae Jo

2025-11-21

Lab report

Exercise 1

```
library(DBI)
library(RSQLite)
library(dbplyr)

con <- dbConnect(SQLite(), "nycflights13.sqlite")
```

Exercise 2

```
library(dbplyr)
flights_tbl <- tbl(con, "flights")
```

Exercise 3

```
flights_query <- flights_tbl %>%
  select(year, month, day, hour, dep_delay, origin)
```

Exercise 4

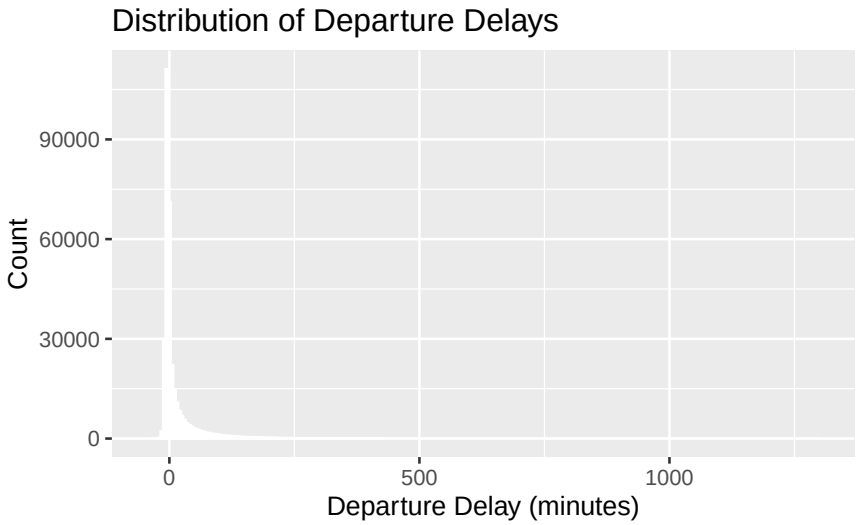
```
## <SQL>
## SELECT `year`, `month`, `day`, `hour`, `dep_delay`, `origin`
## FROM `flights`
```

Exercise 5

```r
flights_df <- flights_query %>%
  collect()
```

Exercise 6

```
## Warning: Removed 8255 rows containing non-finite outside the scale range
## (`stat_bin()`).
```



- The average is much higher than the median, meaning there are many very long delays that drive the average up. While most flights don't experience significant delays, some have very long delays.

Exercise 7

i. Each row represents a weather observation record of a specific date and time at an airport in one of the JFK/LGA/EWRs.

ii. The precip column contains information related to rain or precipitation.

iii. The origin column indicates at which airport (JFK, LGA, EWR) the meteorological measurements were made.

Exercise 8

```
## <SQL>
## SELECT `origin`, `year`, `month`, `day`, `hour`, `temp`, `wind_speed`, `precip`
## FROM `weather`
```

Exercise 9

Exercise 10

Exercise 11

Exercise 12