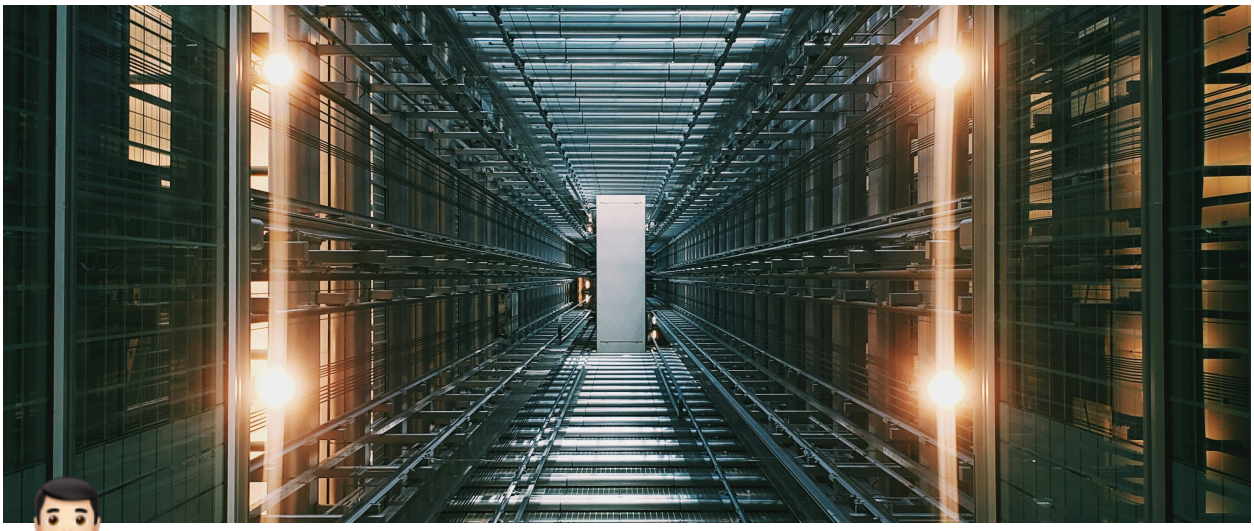

Homework Data Transformation

- HW1: Ask 5 questions about nycflights13 dataframe.

```
## transform nycflights13
library(nycflights13)
library(tidyverse)

data("airlines")
data("airports")
data("flights")
```

  - Q1: Which month has the most flights in 2013?

```
count_month_flights <- flights %>%
  select(year, month) %>%
  filter(year == 2013) %>%
  group_by(month) %>%
```

```
  summarise(count = n()) %>%
  arrange(desc(count))
```

```
# A tibble: 12 × 2
   month count
   <int> <int>
 1     7 29425
 2     8 29327
 3    10 28889
 4     3 28834
 5     5 28796
 6     4 28330
 7     6 28243
 8    12 28135
 9     9 27574
10    11 27268
11     1 27004
12     2 24951
```

▼ Q2: Which airlines has the most top 10 flights?

```
top_flights <- flights %>%
  select(carrier) %>%
  left_join(airlines,by="carrier") %>%
  group_by(name) %>%
  summarise(num_flight = n()) %>%
  arrange(desc(num_flight)) %>%
  head(5)
```

```
# A tibble: 5 × 2
  name                     num_flight
  <chr>                         <int>
1 United Air Lines Inc.         58665
```

```
2 JetBlue Airways                54635
3 ExpressJet Airlines Inc.       54173
4 Delta Air Lines Inc.           48110
5 American Airlines Inc.         32729
```

▼ Q3: Which New York City destinations were the most popular?

```
dest <- flights %>%
  left_join(airports, by = c("dest" = "faa")) %>%
  select(name) %>%
  group_by(name) %>%
  summarise(count_dest = n()) %>%
  arrange(desc(count_dest))
```

```
# A tibble: 102 × 2
   name                                count_dest
   <chr>                                    <int>
 1 Chicago Ohare Intl                       17283
 2 Hartsfield Jackson Atlanta Intl          17215
 3 Los Angeles Intl                         16174
 4 General Edward Lawrence Logan Intl       15508
 5 Orlando Intl                             14082
 6 Charlotte Douglas Intl                   14064
 7 San Francisco Intl                       13331
 8 Fort Lauderdale Hollywood Intl           12055
 9 Miami Intl                               11728
10 Ronald Reagan Washington Natl             9705
```

▼ Q4: Find avg departure delay and avg arrival delay for each airlines?

```
avg_flights <- flights %>%
  filter(arr_delay != "NA" & dep_delay != "NA") %>%
  group_by(carrier) %>%
  summarise(avg_arr_delay = mean(arr_delay),
            avg_dep_delay = mean(dep_delay)) %>%
```

```
  arrange(desc(avg_arr_delay)) %>%
  left_join(airlines, by = "carrier")
```

```
# A tibble: 16 × 4
   carrier avg_arr_delay avg_dep_delay name
   <chr>           <dbl>         <dbl> <chr>
 1 F9             21.9           20.2  Frontier Airline
 2 FL             20.1           18.6  AirTran Airways
 3 EV             15.8           19.8  ExpressJet Airli
 4 YV             15.6           18.9  Mesa Airlines In
 5 OO             11.9           12.6  SkyWest Airlines
 6 MQ             10.8           10.4  Envoy Air
 7 WN              9.65          17.7  Southwest Airlin
 8 B6              9.46          13.0  JetBlue Airways
 9 9E              7.38          16.4  Endeavor Air Inc
10 UA              3.56          12.0  United Air Lines
11 US              2.13           3.74 US Airways Inc.
12 VX              1.76          12.8  Virgin America
13 DL              1.64           9.22 Delta Air Lines
14 AA              0.364          8.57 American Airline
15 HA             -6.92           4.90 Hawaiian Airline
16 AS             -9.93           5.83 Alaska Airlines
```

▼ Q5: Find Average speed for each airlines?

```
# distance in miles
# air_time in minutes

flights %>%
  filter(distance != "NA" & air_time != "NA") %>%
  group_by(carrier) %>%
  mutate(speed = distance / (air_time /60) ) %>%
  summarise(avg_speed = mean(speed)) %>%
```

```
  arrange(desc(avg_speed)) %>%
  left_join(airlines, by = "carrier")
```

```
# A tibble: 16 × 3
   carrier avg_speed name
   <chr>       <dbl> <chr>
 1 HA           480. Hawaiian Airlines Inc.
 2 VX           446. Virgin America
 3 AS           444. Alaska Airlines Inc.
 4 F9           425. Frontier Airlines Inc.
 5 UA           421. United Air Lines Inc.
 6 DL           418. Delta Air Lines Inc.
 7 AA           417. American Airlines Inc.
 8 WN           401. Southwest Airlines Co.
 9 B6           400. JetBlue Airways
10 FL           394. AirTran Airways Corporation
11 MQ           368. Envoy Air
12 OO           366. SkyWest Airlines Inc.
13 EV           363. ExpressJet Airlines Inc.
14 9E           345. Endeavor Air Inc.
15 US           342. US Airways Inc.
16 YV           332. Mesa Airlines Inc.
```

▼ HW2: Create and write a table into Postgresql server (Pizza Restaurant)

```
library(RPostgreSQL)
library(tidyverse)

## Connect to PostgreSQL
con <- dbConnect(
  PostgreSQL(),
  host = "floppy.db.elephantsql.com",
  dbname = "sfouqfvs",
```

```
  user = "sfouqfvs",
  password = "nq5XrYE9g2ANazo42GaWQEPqtDCmn_SB",
  port = 5432
)

## Create a table
customers <- tribble(
  ~customer_id, ~name, ~surname, ~email,
  1, 'David', 'Jones', 'davidjones@gmail.com',
  2, 'Jenny', 'Brown', 'jennybrown@gmail.com',
  3, 'Eric', 'Williams', 'ericwilliams@gmail.com',
  4, 'Jane', 'Hill', 'janehill@gmail.com',
  5, 'Taylor', 'Walker', 'taylorwalker@gmail.com',
  6, 'Paul', 'Turner', 'paulturner@gmail.com',
  7, 'Chris', 'Edwards', 'chrisedwards@gmail.com',
  8, 'Kim', 'Brooks', 'kimbrooks@gmail.com',
  9, 'Mike', 'Hughes', 'mikehughes@gmail.com',
  10, 'Lisa', 'Jackson', 'lisajackson@gmail.com'
)

orders <- tribble(
  ~order_id, ~order_date, ~customer_id, ~menu_id, ~quantity,
  1, '2024-01-26', 3, 12, 1,
  2, '2024-01-26', 6, 14, 1,
  3, '2024-01-27', 8, 11, 2,
  4, '2024-01-27', 4, 13, 3,
  5, '2024-01-27', 9, 15, 1,
  6, '2024-01-28', 5, 7, 1,
  7, '2024-01-28', 10, 5, 2,
  8, '2024-01-28', 1, 10, 1,
  9, '2024-01-28', 7, 9, 2,
  10,'2024-01-29', 9, 8, 1,
  11,'2024-01-29', 6, 4, 2,
  12,'2024-01-29', 3, 3, 1,
  13,'2024-01-29', 4, 1, 2,
  14,'2024-01-30', 2, 6, 3,
```

```
  15,'2024-01-30', 1, 2, 1
)

menus <- tribble(
  ~menu_id, ~menu_name , ~menu_size, ~menu_price, ~category,
  1, 'Pepperoni', 'S', 299, 'Meat',
  2, 'Pepperoni', 'M', 399, 'Meat',
  3, 'Pepperoni', 'L', 499, 'Meat',
  4, 'Hawaiian', 'S', 299, 'Meat',
  5, 'Hawaiian', 'M', 399, 'Meat',
  6, 'Hawaiian', 'L', 499, 'Meat',
  7, 'Seafood', 'S', 399, 'Seafood',
  8, 'Seafood', 'M', 499, 'Seafood',
  9, 'Seafood', 'L', 599, 'Seafood',
  10, 'Meat & Ham', 'S', 299, 'Meat',
  11, 'Meat & Ham', 'M', 399, 'Meat',
  12, 'Meat & Ham', 'L', 499, 'Meat',
  13, 'Mushroom', 'S', 299, 'vegetarian',
  14, 'Mushroom', 'M', 399, 'vegetarian',
  15, 'Mushroom', 'L', 499,'vegetarian'
)


## Write Table to PostgreSQL
dbWriteTable(con, "customers", customers)
dbWriteTable(con, "orders", orders)
dbWriteTable(con, "menus", menus)


## db List Tables
dbListTables(con)


## db List Fields
dbListFields(con, "customers")
```

```
df <- dbGetQuery(con, "select * from customers")


## Disconnect
dbDisconnect(con)
```