HW 01 (dataset "Flights" 5-10 questions)

Q1: In 2013, what season did most people travelled from New York City to other destination?

```
group_ss <- mutate(df_flights,
                   seasonal = case_when(month %in% c(6,7,8) ~ "Summer",
                                        month %in% c(9,10,11) ~ "Autumn",
                                        month %in% c(12,1,2) ~ "Winter",
                                        month %in% c(3,4,5) ~ "Spring"))
group_ss %>% count(seasonal) %>% arrange(desc(n))
```

A spec_tbl_df: 4 × 2

seasonal	n
<chr>	<int>
Summer	86995
Spring	85960
Autumn	83731
Winter	80090

Q2:Which airlines were the most popular in summer 2013?

```
group_ss %>% filter(seasonal == "Summer") %>% count(carrier) %>%
    arrange(desc(n)) %>% left_join(df_airlines,by = "carrier")
```

Q3: Which the top five most popular destinations have been visited in summer 2013?

```
group_ss %>%
    filter(seasonal == "Summer") %>% count(dest) %>% arrange(desc(n)) %>% head(5)
    left_join(df_airports,by = "dest") %>% select(dest,n,"airport name" = name)
```

Q4: What month was frequently delayed?

```
delay_t <- df_flights %>%
filter(!is.na(dep_delay), !is.na(arr_delay)) #ต้องไม่เท่ากับNA

delay_t %>%
  group_by(month) %>%
  summarise(mean = round(mean(dep_delay),2),
              sd = round(sd(dep_delay),2)) %>%
  arrange(desc(mean)) %>%
  head(15)
```

A tibble: 12 × 3

month	mean	sd
<dbl>	<dbl>	<dbl>
7	21.52	51.24
6	20.73	51.29
12	16.48	41.73
4	13.85	42.89
3	13.16	40.05
5	12.89	39.18
8	12.57	37.60
2	10.76	36.17

Q5: Airplanes flew from JFK to LAS, the Turbo Fans more efficient and economical than other types of engines? (JFK to LAS 2248 miles)`

```
planes <- mutate(df_flights,milespermins = distance/((hour*60.0)+minute)) %>%
    left_join(df_planes, by = "tailnum") %>% #select(engine,tailnum,distance,hour
    filter(!is.na(engine),origin == "JFK", dest == "LAS") %>%
    group_by(engine) %>%
    summarise(max_mpm = round(max(milespermins,na.rm = TRUE),2),
              mean_mpm = round(mean(milespermins,na.rm = TRUE),2),
              median_mpm = round(median(milespermins,na.rm = TRUE),2),
              count = n()) %>%
              arrange(desc(count),desc(max_mpm))
planes
```

A tibble: 6 × 5

engine	max_mpm	mean_mpm	median_mpm	count
<chr>	<dbl>	<dbl>	<dbl>	<int>
Turbo-fan	5.87	3.48	3.75	2037
Turbo-jet	4.68	2.64	2.35	1332
Reciprocating	5.76	3.90	3.75	17
Turbo-shaft	3.77	2.62	2.02	8
4 Cycle	5.42	2.83	2.18	5
Turbo-prop	5.42	4.35	5.42	3

HW 02 (3-4 df, RPostgreSQL)

```
#install.packages("RPostgreSQL") Bug
#library(RPostgreSQL) Bug
df_friends <- data.frame(
    id_friends <- 1:4,
    name = c("Brown","Sally","Cony","Choco"),
    fav_color = c("Yellow","Orange","Pink","Purple"),
    age = c(25,20,22,18)
)
df_foods <- data.frame(
    id_menu <- 1:4,
    menu = c("Salmon","Burger","Cake","Brownie"),
    price = c(159,89,99,59)
)
df_tracks <- data.frame(
    tracks_id = c(1,2,5,8),
    tracks_name = c("Magic Man","Pink Venom","Love Story","Gone")
)
```