nycflights13

Nut Th.

2024-05-25

R Markdown

Homework: List up five questions from nycflights13 and use R to find the answer by including a graph.

1) Install package & library

```
  - nycflights13 : Data
  - dplyr        : For data transformation
  - ggplot2      : For create graph
```

```
library(nycflights13)
library(dplyr)
library(ggplot2)
```

2) Downlaod data

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

3) Data Check & Data Preparation

```
#Check NULL
ifelse(mean(complete.cases(flights))==1,"No NA value","Have NA")
```

```
## [1] "Have NA"
```

```
#Drop NULL
flights <- flights[complete.cases(flights),]

# Convert character to factor (For easy to create graph)
flights <- flights %>%
           mutate_if(is.character, as.factor)
```

4) Let's start doing it!

Question 1: How many MIA Delayed Flights are there in Jun to Dec?

- Filter data –> Destination = MIA, Delayed > 0, Month = Jun - Dec

```
MIA_flights <- flights %>%
    filter(dest == 'MIA', dep_delay > 0, month >=6 & month <=12)
```

- Convert month (int) to factor (For graph creation)

```
MIA_flights$month <- factor(
  MIA_flights$month,
  levels = 6:12,
  labels = c('Jun','Jul','Aug','Sep','Oct','Nov','Dec'),
  ordered = TRUE
)
```

- Data transformation & graph creation

```
MIA_flights_chart <- MIA_flights %>%
    count(month) %>%
    ggplot(aes(x = reorder(month, -n), # -n = desc
               y = n,
               fill = month)) +
    geom_col() +
    theme_minimal() +
    scale_fill_brewer(palette = 'Set3') +
    geom_label(aes(label = n), vjust = -0.2, label.size = NA,
                   fill = 'white') +
    labs(title = 'MIA Flight Delays Jun-Dec 2013',
         x = 'Month',
         y = 'Number of Flight Delays',
         fill = 'Flight Delay Months')

MIA_flights_chart
```

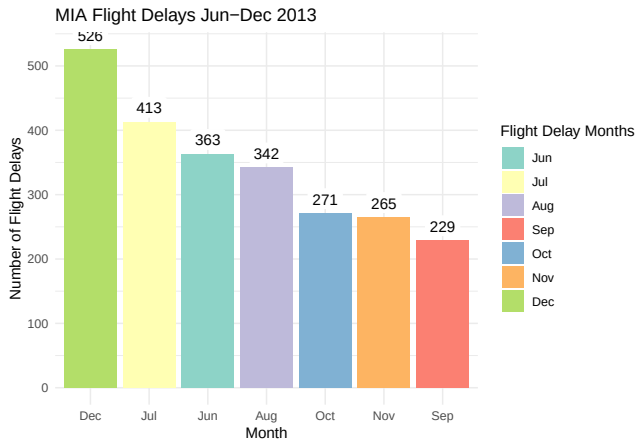

Figure 1: MIA Delayed Flights are there in Jun to Dec

Question 2: What is Top 10 destination in Dec?

- Filter data –> Month = 12

```
flights_dec <- filter(flights, month == 12)
```

- Data transformation

```
(flights_dec_top10 <- flights_dec %>%
                      count(dest) %>%
                      arrange(-n) %>%
                      head(10))
```

```
## # A tibble: 10 x 2
##    dest      n
##    <fct> <int>
##  1 ATL    1429
##  2 LAX    1390
##  3 MCO    1203
##  4 SFO    1159
##  5 CLT    1155
##  6 ORD    1143
##  7 BOS    1096
##  8 MIA    1091
##  9 FLL    1090
## 10 PBI     741
```

- Graph creation

```
flights_dec_chart <- ggplot(flights_dec_top10,
   aes (x = reorder(dest, -n),
        y = n,
        fill = dest)) +
   geom_bar(stat = 'identity') +
   theme_minimal()+
   scale_fill_brewer(palette = "PRGn") +
   geom_label(aes(label = n), vjust = -0.2, label.size = NA,
                  fill = 'white') +
   labs(title = 'The Most Popular Destination in Dec 2013',
        x = 'Destination',
        y = 'Count')

flights_dec_chart
```

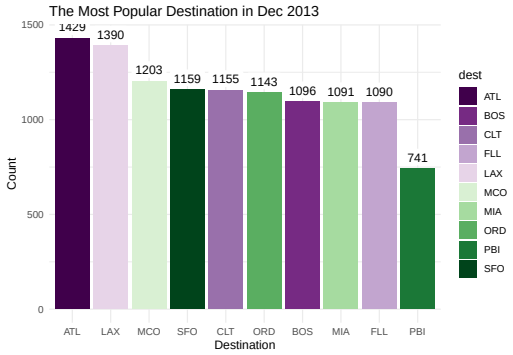

Figure 2: Top 10 destination in Dec

Question 3: What is an average monthly flight departing from JLK?

- Filter data –> Origin = 'JLK'. Select Col –> month, day, origin, dest, tailnum)

```
JFK_flights <- flights %>%
               filter(origin == 'JFK') %>%
               select(month, day, origin, dest, tailnum)
```

- Data transformation

```
(JFK_flights_avg_monthly <- JFK_flights %>%
                            group_by(month) %>%
                            summarise(avg_flights = mean(n())))
```

```
## # A tibble: 12 x 2
##    month avg_flights
##    <int>       <dbl>
##  1     1        9031
##  2     2        8007
##  3     3        9497
##  4     4        9013
##  5     5        9270
##  6     6        9182
##  7     7        9757
##  8     8        9870
##  9     9        8788
## 10    10        9096
## 11    11        8645
## 12    12        8923
```

- Convert month to factor

```
JFK_flights_avg_monthly$month <- factor(JFK_flights_avg_monthly$month,
    levels = 1:12,
    labels = c('Jan','Feb','Mar','Apr',
               'May','Jun','Jul','Aug',
               'Sep','Oct','Nov','Dec'),
    ordered = TRUE)
```

- Graph creation

```
JFK_flights_chart <- ggplot(JFK_flights_avg_monthly,
                     aes(x = month,
                         y = avg_flights,
                         fill = month))+
                     geom_bar(stat = 'identity') +
                     theme_minimal() +
                     coord_flip() +
                     scale_fill_brewer(palette = 'Set3') +
                     geom_label(aes(label = avg_flights), hjust = 0,
                                label.size = NA, fill = 'white') +
                     labs(title = 'Average Monthly Flights departing from JFK in 2019',
                         x = 'Month',
                         y = 'Number of Flights')

JFK_flights_chart
```

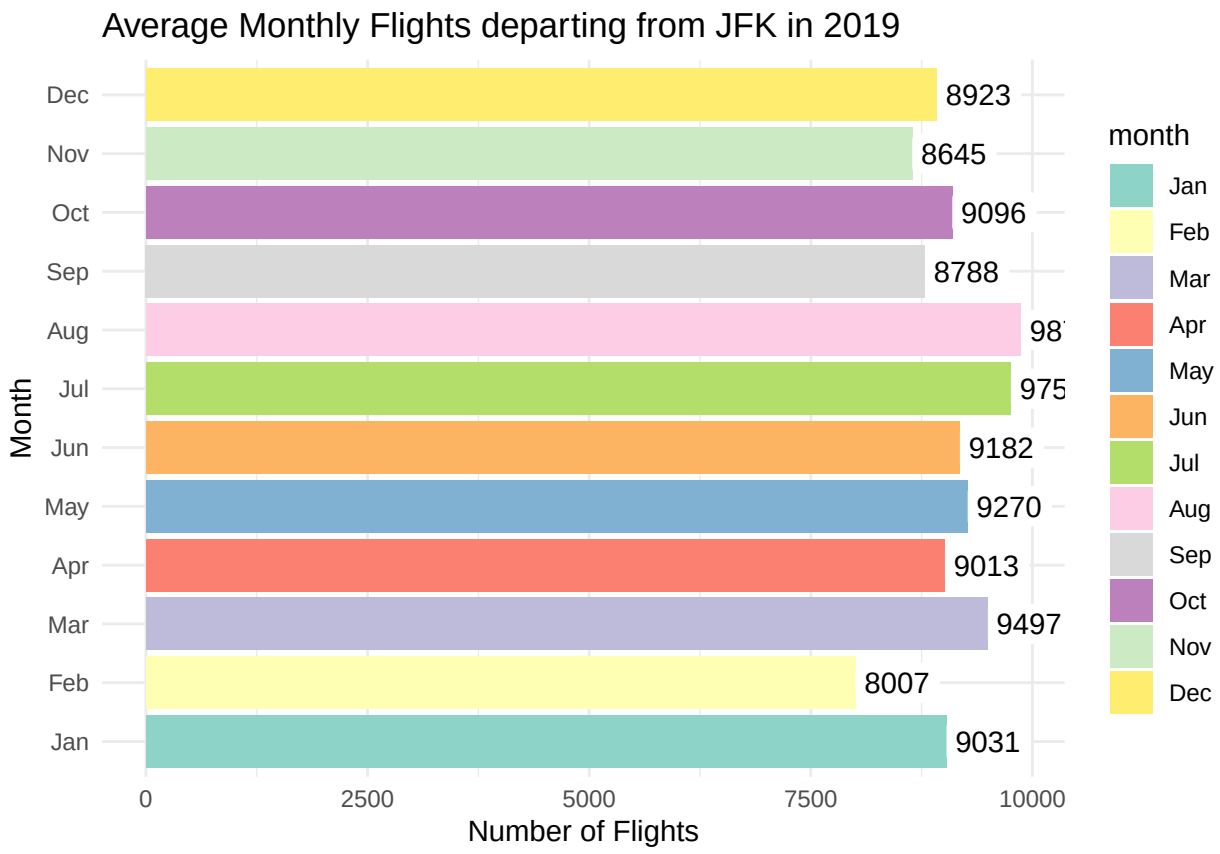

Figure 3: Average monthly flight departing from JLK