STAT409 Homework #1

2021320322 / Minseo Yoon

March 21, 2023

(a)

```
library(nycflights13)
library(tidyverse)

flights <- nycflights13::flights
flights %>%
  group_by(month) %>%
  summarize(cancelled_proportion =
              sum(is.na(dep_time) | is.na(arr_time)) / n()) %>%
  arrange(desc(cancelled_proportion))
```

```
## # A tibble: 12 x 2
##    month cancelled_proportion
##    <int>                <dbl>
##  1     2              0.0517
##  2     6              0.0380
##  3    12              0.0376
##  4     7              0.0354
##  5     3              0.0309
##  6     4              0.0251
##  7     5              0.0209
##  8     1              0.0198
##  9     9              0.0183
## 10     8              0.0173
## 11    11              0.00928
## 12    10              0.00855
```

It was canceled at the largest rate in February and the smallest rate in October.

December and February seem to be at the top of the list due to winter snow.

Similarly, many flights appear to be at the top of the list, with June and July canceled due to summer rain.

(b)

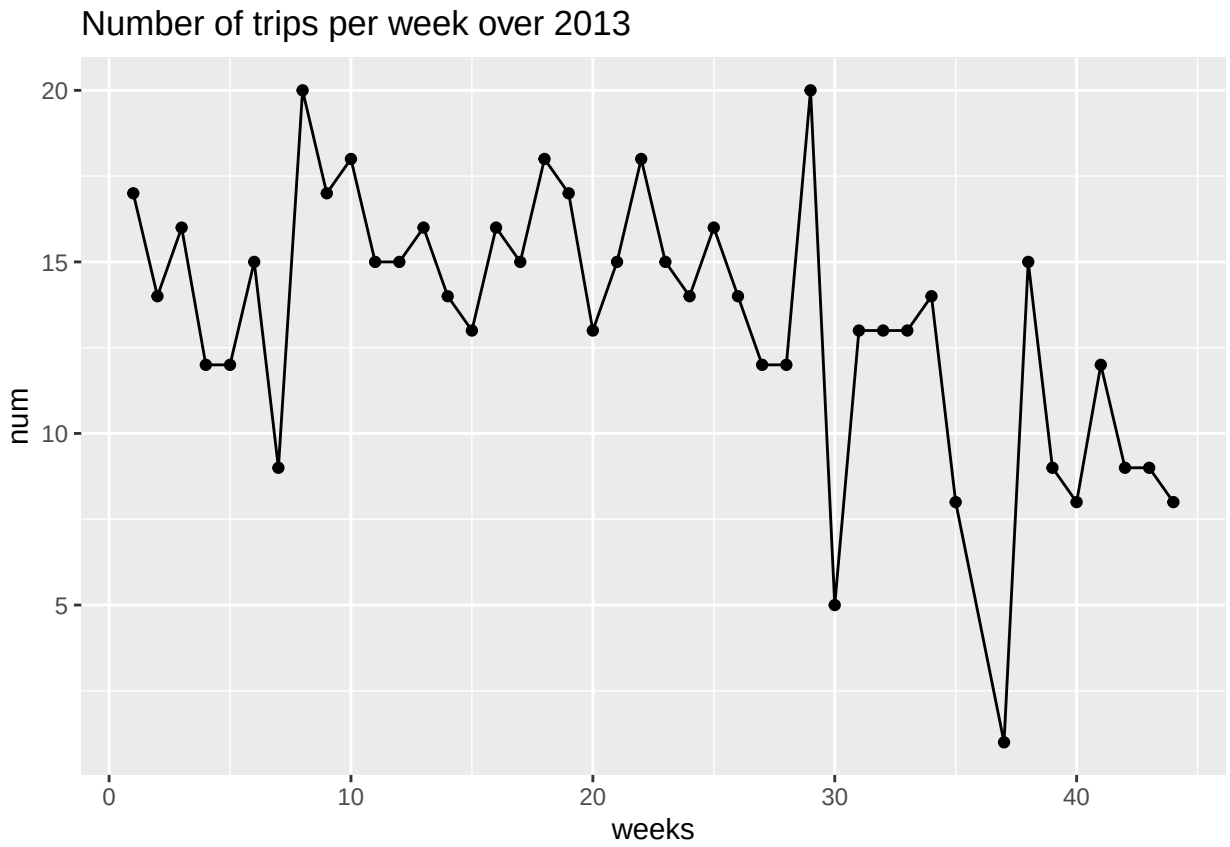
```
library(lubridate)

flights %>%
  filter(!is.na(tailnum)) %>%
  group_by(tailnum) %>%
  summarize(num = n()) %>%
  arrange(desc(num)) %>%
  head(1)
```

```
## # A tibble: 1 x 2
##   tailnum   num
##   <chr>   <int>
## 1 N725MQ    575
```

The N725MQ flew the most with 575.

```
flights %>%
  filter(tailnum == 'N725MQ') %>%
  mutate(date = paste(sprintf('%04d-%02d-%02d', year, month, day)),
         weeks = week(date)) %>%
  group_by(weeks) %>%
  summarize(num = n()) %>%
  ggplot(aes(x = weeks, y = num)) +
  geom_point() +
  geom_line() +
  ggtitle('Number of trips per week over 2013')
```


Number of trips per week over 2013
num
20
15
10
5
0
10
20
30
40
weeks

2.

$g(a) = E(X^2 - 2aX + a^2)$

$= a^2 - 2aE(X) + E(X^2)$ ⟶ a에 대한 이차함수

convex이므로 $g'(a)=0$인 a가 minimizer

$g'(a) = 2a - 2E(X) \overset{set}{=} 0$

$\therefore a = E(X)$

3.

$P(|\bar{X}_n - E(X)| > \varepsilon)$에서 Chebyshev inequality에 의해

$$P(|\bar{X}_n - E(X)| > \varepsilon) \le \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$$

4.

(a)

$P_A P_A = A(A^TA)^{-1}A^TA(A^TA)^{-1}A^T$

$= A(A^TA)^{-1}A^T = P_A$

(b)

$P_A z = P_A P_A v = P_A v = z$

(c)

$(v-z)^T(u-z) = v^Tu - v^Tz - z^Tu + z^Tz$

$= v^TP_Au - v^TP_Av - v^TP_A^Tu + v^TP_A^TP_Av$ ↙ ($\because P_A$: idempotent, symmetric)

$= v^TP_Au - v^TP_Av - v^TP_Au + v^TP_Av$

$= 0$

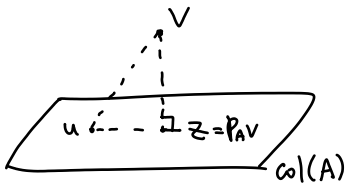
$\therefore (v-z)$과 $(u-z)$는 수직이다.

↘ $P_A^T = \{A(A^TA)^{-1}A^T\}^T$

$= A\{(A^TA)^T\}^{-1}A^T$

$= A(A^TA)^{-1}A^T = P_A$

따라서 오른쪽 그림에서 피타고라스 정리에 의해서

$\|v-u\|^2 = \|v-z\|^2 + \|u-z\|^2$

$\ge \|v-z\|^2$ ($\because \|u-z\|^2 \ge 0$)