326.212 Final Project: Part 3

2018-11120 양은우

Last Modified 8th Nov, 12:34pm (KST)

Part 3. Advanced Questions (100 pts)

Provide a graphical summary to answer the following questions. These are intentionally vague in order to allow you to focus on different aspects of the data. We only provide some suggestions.

Q1. When is the best time of day/day of week/time of year to fly to minimize delays? (50 pts)

Suggestions: consider flights the not arrives delayed. You may want to ignore negative delays.

Distribution of scheduled departure time and delays

```
delays_clean %>%
  ggplot(aes(x = sched_dep_time, y = dep_delay)) +
  geom_hex(binwidth = c(60, 30)) +
  scale_fill_viridis_c() +
  theme_bw()
```

```
delays_departure %>%
  ggplot(aes(x = sched_dep_time, y = dep_delay)) +
  geom_hex(binwidth = c(60, 30)) +
  scale_fill_viridis_c() +
  theme_bw()
```

```
airlines_dep_time = delays_clean %>%
  filter(carrier%in%c) %>%
  ggplot(aes(x = sched_dep_time, y = dep_delay, fill = carrier)) +
  geom_hex(binwidth = c(60, 30)) +
  facet_wrap(~ carrier, nrow = 3) +
  scale_fill_viridis_c(option = "inferno", guide = FALSE) +
  theme_bw()
airlines_dep_time
```

```
airlines_dep_time = delays_departure %>%
  filter(carrier%in%c) %>%
  ggplot(aes(x = sched_dep_time, y = dep_delay, fill = carrier)) +
  geom_hex(binwidth = c(60, 30)) +
  facet_wrap(~ carrier, nrow = 3) +
  scale_fill_viridis_c(option = "inferno", guide = FALSE) +
  theme_bw()
airlines_dep_time
```

위의 두 그래프를 보면 대부분의 지연은 오전 시간에 발생하는 것을 알 수 있다. 이는 아마도 공항 혼잡도 때문에 오전 시간에 많은 비행기들이 출발하기 때문에 발생하는 지연이라고 생각할 수 있다. 또한 밤에 출발하는 경우도 늦어지는 경우가 조금 있는데 이는 아마도 기상 상태 때문이라고 생각할 수 있다. 또한 일반적으로 출발하는 시간이 늦어지면 지연되는 경우 많음을 알 수 있다.

The best time of day to fly to minimize delays

```
delays_clean %>%
  ggplot(aligned_dep_time, dep_delay) +
  geom_hex(binwidth = c(60, 30)) +
  facet_wrap(~ carrier, nrow = 3) +
  scale_fill_viridis_c(guide = FALSE) +
  theme_bw()
```

```
delays_clean %>%
  ggplot(aligned_dep_time, dep_delay) +
  geom_hex(binwidth = c(60, 30)) +
  facet_wrap(~ carrier, nrow = 3) +
  scale_fill_viridis_c(guide = FALSE) +
  theme_bw()
```

```
ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

위 그래프 들 가운데 가장 좋은 시간대를 고르는 것을 선택하기 위해 Delta Airlines 에서는 대부분 지연되는 것을 알 수 있다. 이러한 경향은 Aligned departure time 이 0 에서 1000 사이 일 때 가장 심하다. 이러한 시간대는 대부분 오전 시간이다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피하는 것이 가장 좋은 방법이다. 또한 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 Aligned departure time 이 1000 이상일 때 비행기를 타는 것이 가장 좋다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 오후 시간에 비행기를 타는 것이 가장 좋다.

The best day of week to fly to minimize delays

```
delays_clean %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_clean %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_departure %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_departure %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

일주일 중 주말을 제외 한 요일 중 Delta Airlines 에서 가장 지연되는 경우가 많다. 이러한 경향은 Aligned departure time 이 0 에서 1000 사이 일 때 가장 심하다. 이러한 시간대는 대부분 오전 시간이다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피하는 것이 가장 좋은 방법이다. 또한 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 Aligned departure time 이 1000 이상일 때 비행기를 타는 것이 가장 좋다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 오후 시간에 비행기를 타는 것이 가장 좋다.

The best time of year to fly to minimize delays

```
delays_clean %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_clean %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_departure %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_departure %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

일주일 중 주말을 제외 한 요일 중 Delta Airlines 에서 가장 지연되는 경우가 많다. 이러한 경향은 Aligned departure time 이 0 에서 1000 사이 일 때 가장 심하다. 이러한 시간대는 대부분 오전 시간이다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피하는 것이 가장 좋은 방법이다. 또한 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 Aligned departure time 이 1000 이상일 때 비행기를 타는 것이 가장 좋다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 오후 시간에 비행기를 타는 것이 가장 좋다.

Q2. Do older planes suffer more delays? (50 pts)

Suggestions: you may want to find a correlation between the age of the plane and the departure delay. If this data size is too big to compute the correlation, you may want to sample a random fraction from the dataset.

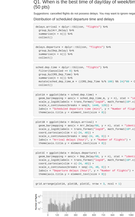
```
delays_clean %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_clean %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_departure %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```

```
delays_departure %>%
  filter(carrier%in%c) %>%
  ggplot(aligned_dep_time, dep_delay, color = carrier) +
  geom_smooth(se = FALSE) +
  scale_color_viridis(discrete = T, guide = "legend") +
  theme_bw()
```


위의 그래프 들 가운데 가장 좋은 시간대를 고르는 것을 선택하기 위해 Delta Airlines 에서는 대부분 지연되는 것을 알 수 있다. 이러한 경향은 Aligned departure time 이 0 에서 1000 사이 일 때 가장 심하다. 이러한 시간대는 대부분 오전 시간이다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피하는 것이 가장 좋은 방법이다. 또한 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 Aligned departure time 이 1000 이상일 때 비행기를 타는 것이 가장 좋다. 이러한 경향은 Delta Airlines 에서 가장 심하다. 따라서 Delta Airlines 에서 비행기를 타는 것을 피할 수 없다면 오후 시간에 비행기를 타는 것이 가장 좋다.



일주일 중 주말을 제외 한 요일 중 Delta Airlines 에서 가장 지연되는 경우가 많