Stat. 450 Quiz preparation

Stat. 450: Quiz preparation

These question is related to the homework problem 5.6.7 Exercise 1.

For the flights data, in the nycflights13 package, answer the following questions:

```
library(nycflights13)
library(tidyverse)
```

1. How many flights arrived on time (which includes the flights that arrived early)?

To answer this question we will look at the arr_delay to look at the flights that arrived on time.

Answer: 194,342

```
flights %>% select(arr_delay) %>%
  filter(arr_delay <= 0) %>%
  summarize( n=n() )
```

```
## # A tibble: 1 x 1
##        n
##    <int>
## 1 194342
```

2. What proportion of flights arrived on time?

To answer this question we will look at the arr_delay to look at the flights that arrived on time.

Answer: Approximately 60%.

```
flights %>% select(arr_delay) %>%
  summarize( arr_delay_mean = mean(arr_delay <= 0, na.rm = TRUE) )
```

```
## # A tibble: 1 x 1
##   arr_delay_mean
##            <dbl>
## 1          0.594
```

3. How many United flights arrived 30 or more minutes late?

Answer: The number of UA flights that arrived 30 or more minutes late was 8131.

```
flights %>% filter( carrier == "UA") %>%
  count(arr_delay >= 30)
```

```
## # A tibble: 3 x 2
##   `arr_delay >= 30`     n
##   <lgl>             <int>
## 1 FALSE             49651
## 2 TRUE               8131
## 3 NA                  883
```

4. Which airline has the best on-time performance?

Answer: AS

```
flights %>% select(arr_delay, carrier) %>%
  group_by(carrier)  %>%
```

```
  summarize( n=n(), arr_delay_mean=mean(arr_delay <= 0, na.rm = TRUE) ) %>%
  arrange(desc(arr_delay_mean))
```

```
## # A tibble: 16 x 3
##    carrier     n arr_delay_mean
##    <chr>   <int>          <dbl>
##  1 AS        714          0.733
##  2 HA        342          0.716
##  3 AA      32729          0.665
##  4 VX       5162          0.659
##  5 DL      48110          0.656
##  6 OO         32          0.655
##  7 US      20536          0.629
##  8 9E      18460          0.616
##  9 UA      58665          0.615
## 10 B6      54635          0.563
## 11 WN      12275          0.560
## 12 MQ      26397          0.533
## 13 YV        601          0.526
## 14 EV      54173          0.521
## 15 F9        685          0.424
## 16 FL       3260          0.403
```

5. Which airline has the worst on-time performance?

Answer: FL

```
flights %>% select(arr_delay, carrier) %>%
  group_by(carrier)  %>%
  summarize( n=n(), arr_delay_mean=mean(arr_delay <= 0, na.rm = TRUE) ) %>%
  arrange(arr_delay_mean)
```

```
## # A tibble: 16 x 3
##    carrier     n arr_delay_mean
##    <chr>   <int>          <dbl>
##  1 FL       3260          0.403
##  2 F9        685          0.424
##  3 EV      54173          0.521
##  4 YV        601          0.526
##  5 MQ      26397          0.533
##  6 WN      12275          0.560
##  7 B6      54635          0.563
##  8 UA      58665          0.615
##  9 9E      18460          0.616
## 10 US      20536          0.629
## 11 OO         32          0.655
## 12 DL      48110          0.656
## 13 VX       5162          0.659
## 14 AA      32729          0.665
## 15 HA        342          0.716
## 16 AS        714          0.733
```