Hebert_M2A

2024-09-08

R Data Visualization and Wrangling Analysis

Titanic Dataset

Question: How did survival among adult passengers vary by sex and passenger class?

Table Summaries

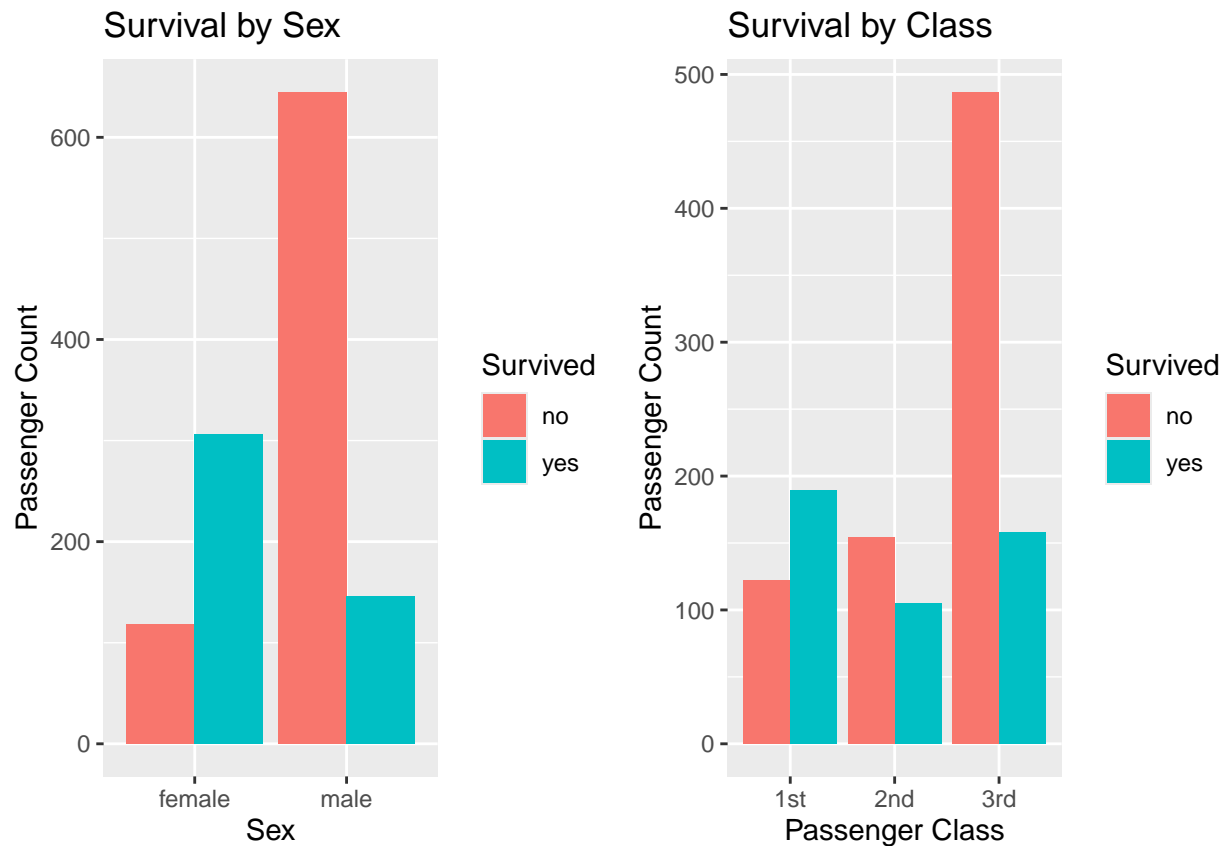Here are the numerical tables with survival between sex and class.

```
## Survival by Sex:

## # A tibble: 4 x 3
##    sex     survived count
##    <chr>   <chr>    <int>
## 1 female no         118
## 2 female yes        306
## 3 male   no         645
## 4 male   yes        146


## Survival by Passenger Class:

## # A tibble: 6 x 3
##   passengerClass survived count
##   <chr>          <chr>    <int>
## 1 1st            no         122
## 2 1st            yes        189
## 3 2nd            no         154
## 4 2nd            yes        105
## 5 3rd            no         487
## 6 3rd            yes        158
```

Bar Graphs of Distribution

It is clear from these charts that a much higher distribution of adult female passengers survived the crash than males. This suggests that women were prioritized in the rescue and survival efforts.

In regards to Passenger Class, the ratio of surviving passengers to non-surviving passengers is not too different for 1st and 2nd class, but the number of non-surviving passengers greatly increases for passengers in 3rd class. Perhaps passengers in higher classes had better resources, life-saving equipment, or priority on life boats compared to those in lower classes.

The last table for this dataset to include is survival grouped by sex and class for comparison.

```
## # A tibble: 12 x 4
##    sex    passengerClass survived count
##    <chr>  <chr>          <chr>    <int>
##  1 female 1st            no           5
##  2 female 1st            yes        132
##  3 female 2nd            no          12
##  4 female 2nd            yes         85
##  5 female 3rd            no         101
##  6 female 3rd            yes         89
##  7 male   1st            no         117
##  8 male   1st            yes         57
##  9 male   2nd            no         142
## 10 male   2nd            yes         20
## 11 male   3rd            no         386
## 12 male   3rd            yes         69
```

Female passengers in 1st and 2nd class had a high rate of survival, whereas women in 3rd class did not. Male passengers of all classes did not have a high rate of survival.

NYC Flights Data Set

Questions:

- What times of year have the worst and best average departure delays? Does this vary by airport?
- Which routes from the NYC area gained the most time in air, on average?
- Which carriers have the best "on-time" performance" (defined as the proportion of flights arriving within 10 minutes of their scheduled arrival time)

Time of Year analysis

```
## Warning: NAs introduced by coercion
```


Delay by Month and Origin Airport
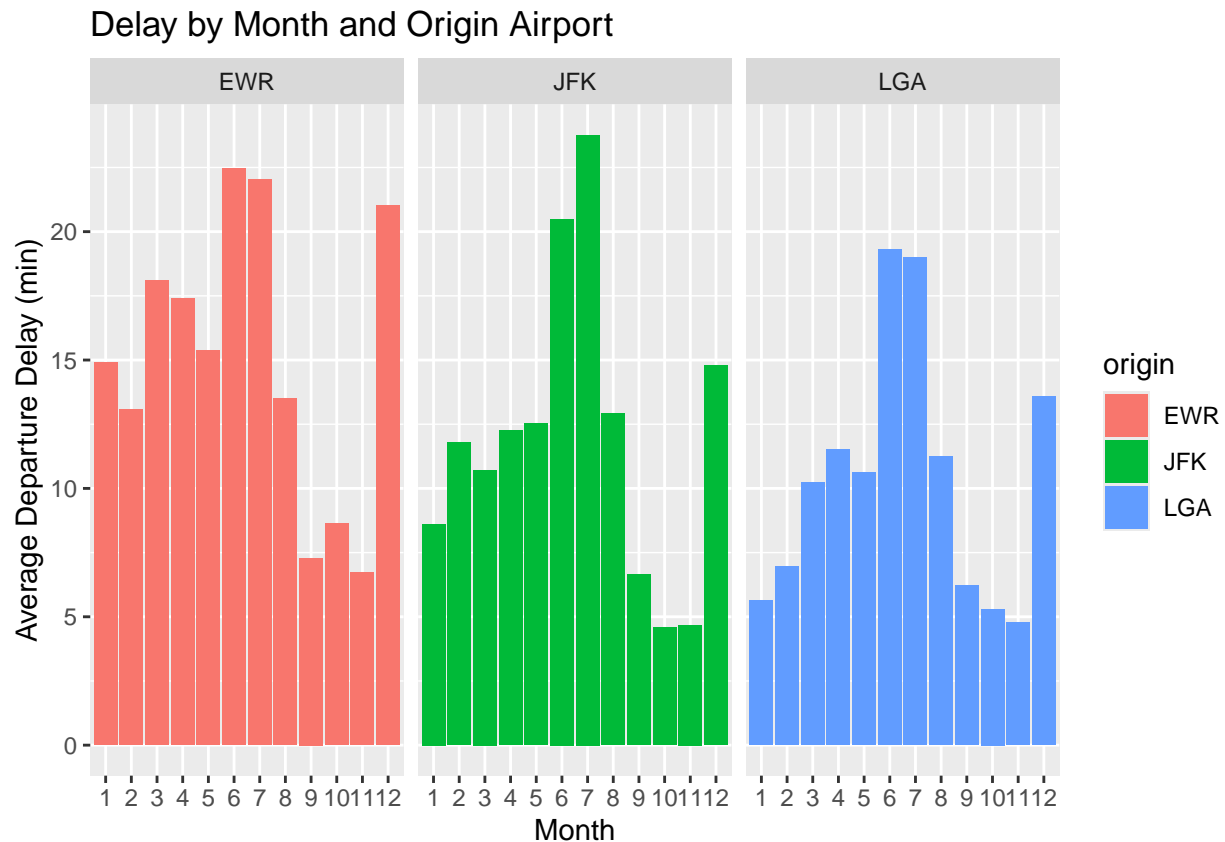
For all 3 airports in the NYC region, delays seem to be highest in the summer months, with an additional spike in December. These are likely the times with the highest number of flights for travel, visiting family, etc. December delays could also be caused by weather.

LGA has lower average delays overall, with EWR and JFK having higher average delays throughout the year. The spikes at EWR and JFK are higher in the summer months as well.

Routes with Best Gain

Warning: NAs introduced by coercion
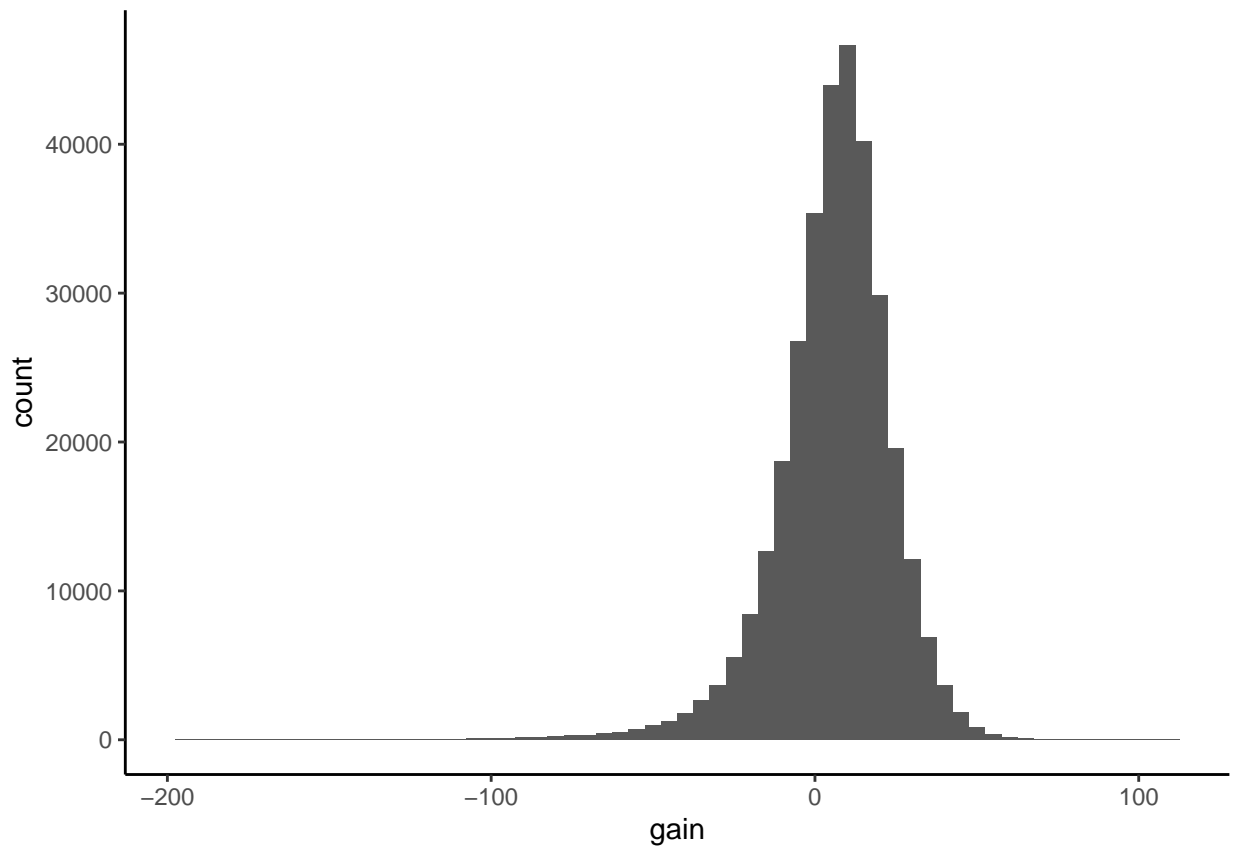
This histogram shows the distribution of gain for flights all across the dataset. Gain is a measure that takes departure delay and subtracts arrival delay. If a flight is 10 minutes late to depart, but arrives only 5 minutes behind the original schedule, the flight made up 5 minutes in the air. Now I will break this out by route to find the "best" ones in terms of gain.

```
## # A tibble: 105 x 2
##    dest   mean_gain
##    <chr>      <dbl>
##  1 MTJ         15.9
##  2 ANC         15.4
##  3 SNA         14.6
##  4 SBN         14.6
##  5 ILM         14.5
##  6 SAT         13.4
##  7 LEX         13
##  8 BHM         12.1
##  9 CHO         11.9
## 10 SEA         11.7
## # i 95 more rows
```

Flights to Montrose, CO show the best gain from NYC airports, with an average of 15.9 on this route.

On-Time Performance

```
## # A tibble: 16 x 4
##    carrier total_flights ot_flights percent_ot
##    <chr>           <int>      <dbl>      <dbl>
##  1 HA                342        283       82.7
##  2 AS                714        578       81.0
##  3 VX               5162       3993       77.4
##  4 DL              48110      37060       77.0
##  5 AA              32729      24781       75.7
##  6 US              20536      15382       74.9
##  7 UA              58665      42721       72.8
##  8 WN              12275       8491       69.2
##  9 OO                 32         22       68.8
## 10 B6              54635      37521       68.7
## 11 9E              18460      12415       67.3
## 12 MQ              26397      16957       64.2
## 13 EV              54173      32906       60.7
## 14 FL               3260       1882       57.7
## 15 YV                601        346       57.6
## 16 F9                685        387       56.5
```

This table shows the % of flights that are on-time for each carrier in the dataset. The following graph shows
the ranked carriers in chart form.
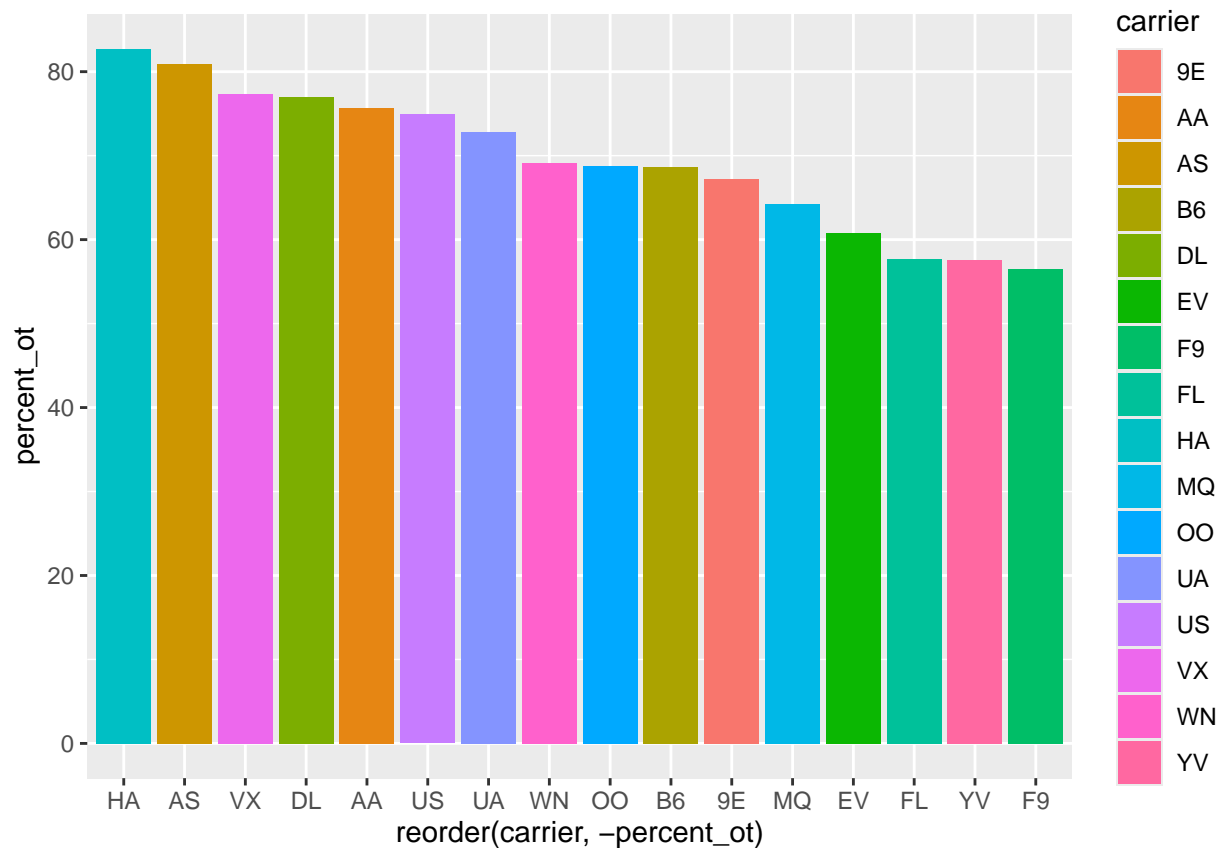
HA, or Hawaiian Airlines, had the best on-time performance in the dataset. Nearly 83% of their flights
arrived within 10 minutes of their scheduled time. Their volume of flights was on the lower end, but this is

still impressive performance. Alaska airlines came in second.