

Lab_06

Fantastic Four

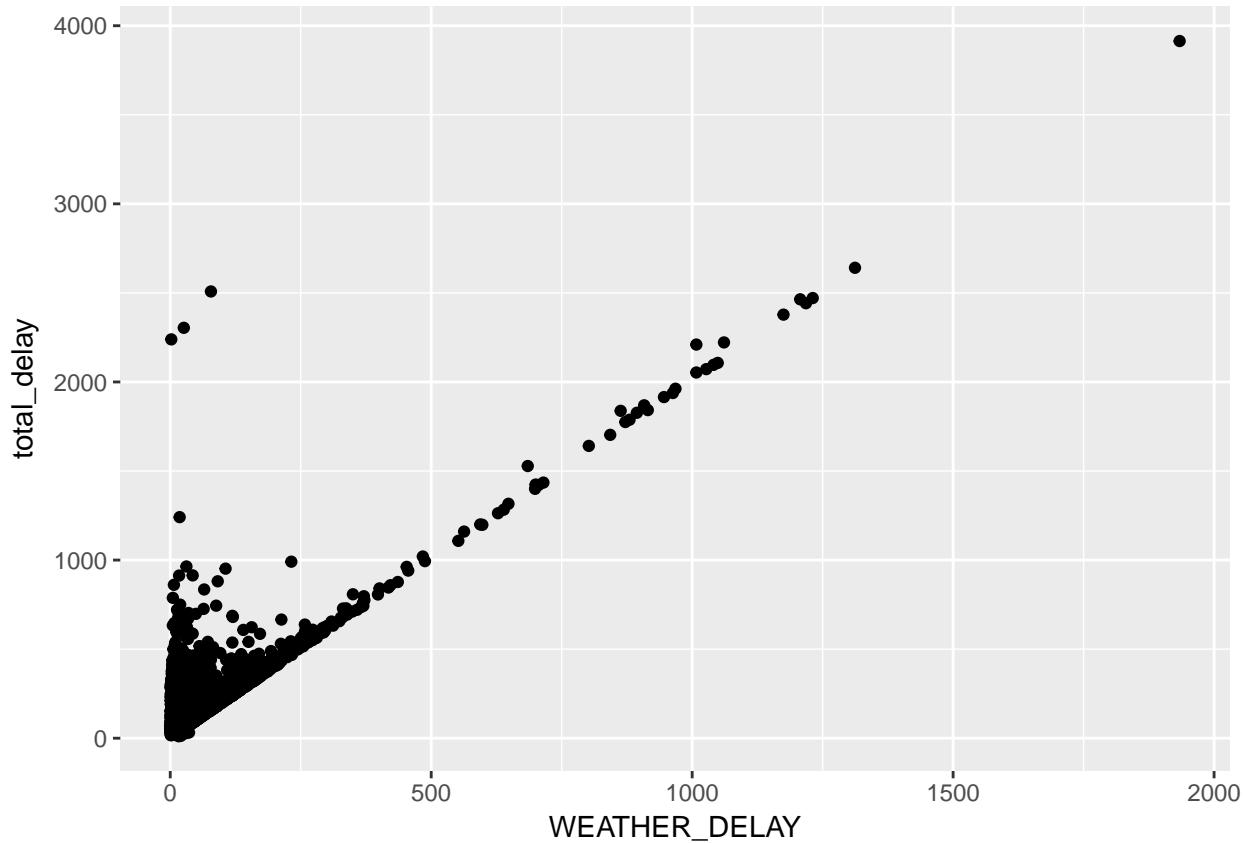
Guiding Question

What factors lead to on time arrivals? What factors cause flight delays?

It is important to isolate the different events which result in delayed flights and prompt flights in order to improve procedures to make them more effective and give a better assurance that flights will remain on schedule. This information could also help schedule future flights as well as preserve all assets which go into operating airports and flights. The factors we examined were: time, airline, weather, flight time, and length of flight. Although it can be difficult to completely isolate each of these factors (as, in the real world factors are compound), we tried to isolate these elements the best we could.

Conclusion

This graph best shows that the main factor of the longest flight delays are caused by weather. The chart below contains the top 10 longest delays in 2017, most of which occur in December through February. This information goes along with what we have found below in the individual findings section about the months relating to amount of delays. There are still points that lie outside of the linear trend which tells us that weather is not the only factor contributing to the delay time. Furthermore, there seems to be a strong correlation (.957) between the departure delay of a flight and the tardiness at DIA—which means that DIA might not be the problem with late flights.



```
## # A tibble: 10 x 3
##   MONTH DAY_OF_MONTH total_delay
##   <int>     <int>      <int>
## 1 1         1             5       3914
## 2 2         1            23      2508
## 3 3         2            11      2304
## 4 4         2            11      2378
## 5 5         2            19      2222
## 6 6         3             4      2641
## 7 7         8            28      2471
## 8 8        12            23      2239
## 9 9        12            21      2464
## 10 10        12            25      2442
```

Overall Findings

There seems to be many factors which all contribute to delayed flights. There are the measured delays such as weather, security, mechanics, and air traffic. There is also a correlation between the month of the year and the amount of flight delays. January seems to have the most delays, and then the delays decrease until April which then begin to increase until August, followed by increasing delays until December. There is also relevant that the origin and destination of flights impacted the delays, for example, flights from Denver are more likely to be delayed than flights into Denver. There is also a relationship between flight length and how much the flight is delayed. On average two hour flights had the most amount of delays. Additionally, JetBlue and Virgin Airlines had the worst proportion of late flights (~27% and ~25% respectively) into Denver.

Recommendations

Dear DIA Administraion, The following observations were made through our analysis:

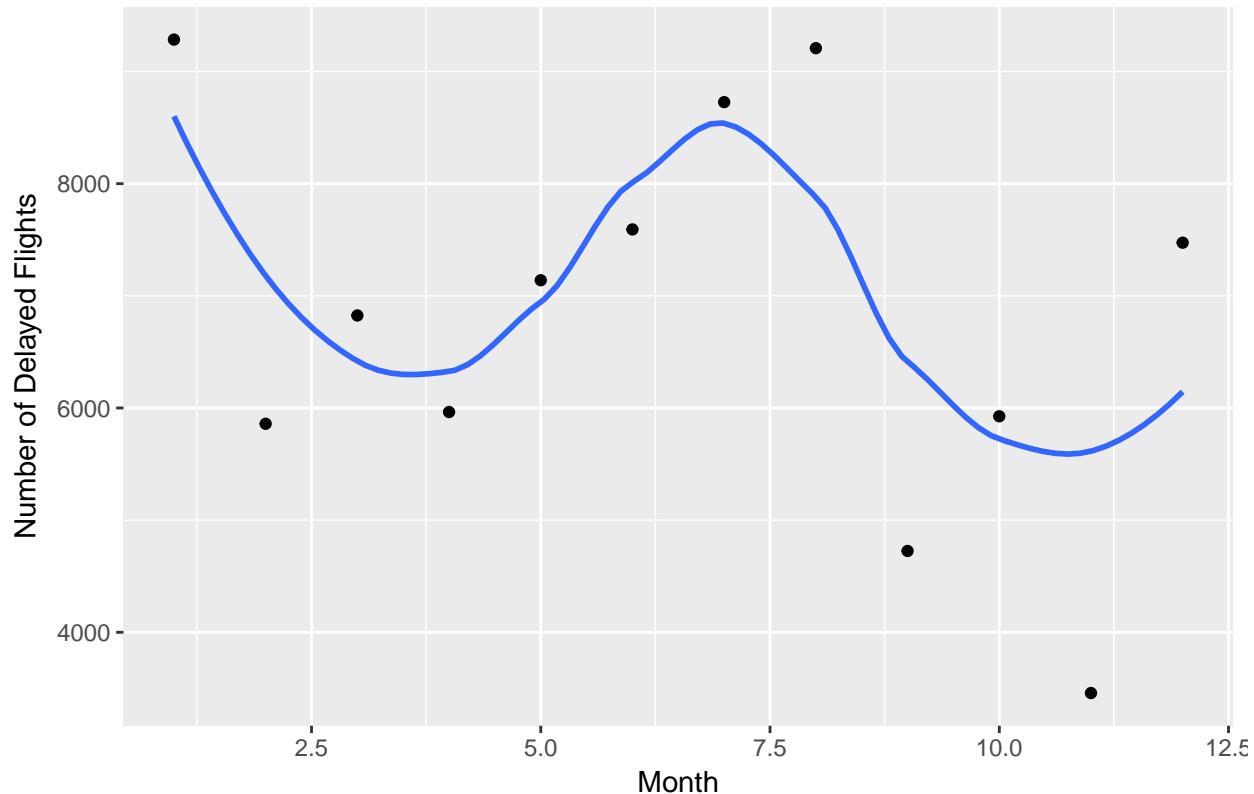
- Weather is not controlable but is a large factor in flight delays—helps to be more prepared in cases of snow, etc. to make sure you minimize financial loss in operations.
- Effective crowd management could improve your delay time at busy times of the year because there seems to be more late flights at those times (summer and the holidays)
- Solve the domino effect by controlling departure delays at the origin, therefore preventing more arrival delays.
- Examine your contracts (so you do not lose money) with JetBlue, Virgin, and Spirit since they have the highest proportions of late flights.
- Think about examining your contracts with Southwest and United since they have a large mass of late flights.

Individual Findings

Lexie Marinelli

```
## `geom_smooth()` using method = 'loess'
```

Trend of delayed flights throughout the year



This graph shows the trend of the number of flights delayed in each month. January tops the other months with the number of flights delayed, it drops down to lower values until April. After April there is close to a linear increase in the number of delayed flights going until August, due to summer vacationers. September drops down significantly, most likely it is not a popular month to travel, as well as November since it has the

lowest number of delayed flights. December increases in their number of delayed flights due to the hectic holiday season travellers.

Lindsay Gettel

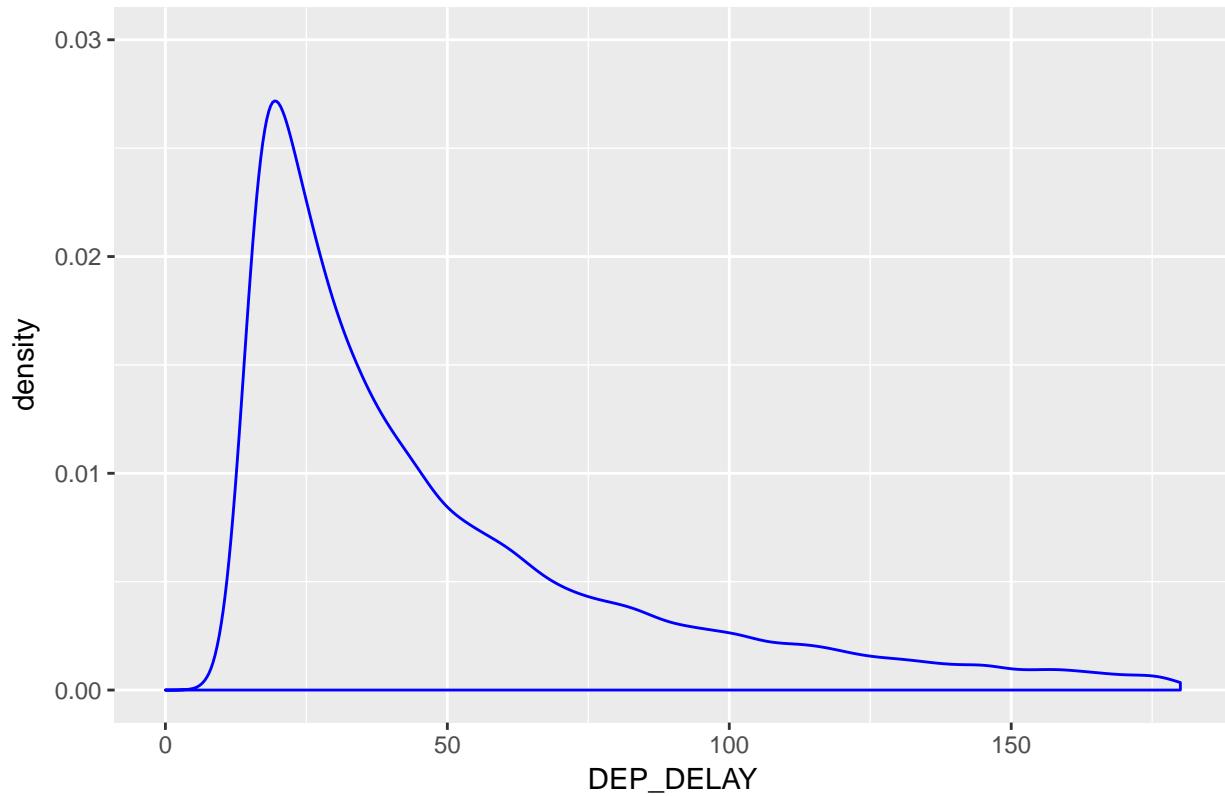
```
#Is a flight more likely to be delayed when leaving DEN or arriving in DEN?  
Coflights%>%
```

```
filter(ORIGIN=="DEN", DEP_DELAY>=15, !CANCELLED=="1")%>%
```

```
ggplot() + geom_density(mapping=aes(x=DEP_DELAY), color="blue") + xlim(0, 180) + ggttitle("Departure delay for flight leaving DEN")
```

```
## Warning: Removed 1402 rows containing non-finite values (stat_density).
```

Departure delay for flight leaving DEN



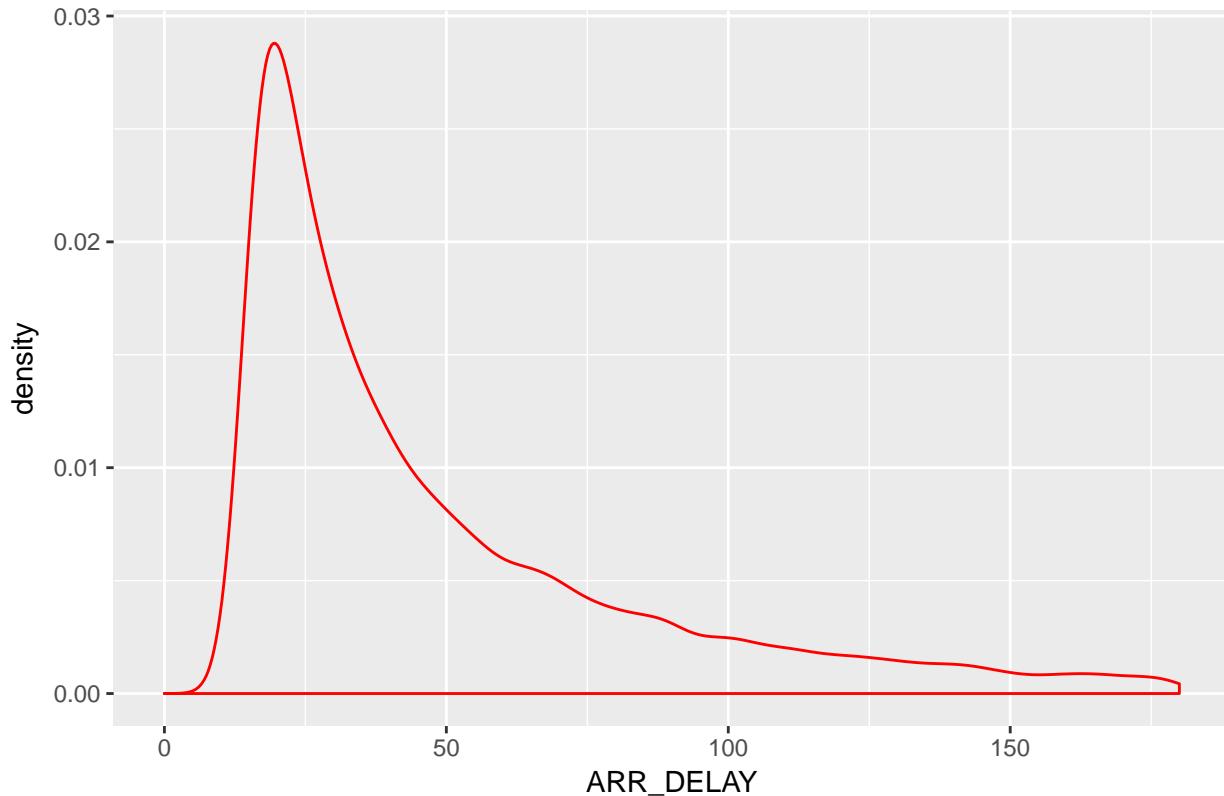
```
Coflights%>%
```

```
filter(DEST=="DEN", ARR_DELAY>=15, !CANCELLED=="1")%>%
```

```
ggplot() + geom_density(mapping=aes(x=ARR_DELAY), color="red") + xlim(0,180) + ggttitle("Arrival delay for flights arriving in DEN")
```

```
## Warning: Removed 1843 rows containing non-finite values (stat_density).
```

Arrival delay for flight into DEN



```
#probability late given the destination was DEN\
allFlights<-filter(COflights, !CANCELLED=="1" & !is.na(CANCELLED))
lateDen<-allFlights%>% filter(DEST=="DEN" & ARR_DELAY>=15)%>% count()
den<-allFlights%>% filter(DEST=="DEN")%>% count()
P_late_toDEN<-lateDen/den
#probability late given origin was DEN
late_from_DEN<-allFlights%>% filter(ORIGIN=="DEN" & ARR_DELAY>=15)%>% count()
fromDen<-allFlights%>% filter(ORIGIN=="DEN")%>% count()
P_late_fromDen<- late_from_DEN/fromDen
message("Conditional probability a flight is late given the destination is DEN: ", P_late_toDEN)

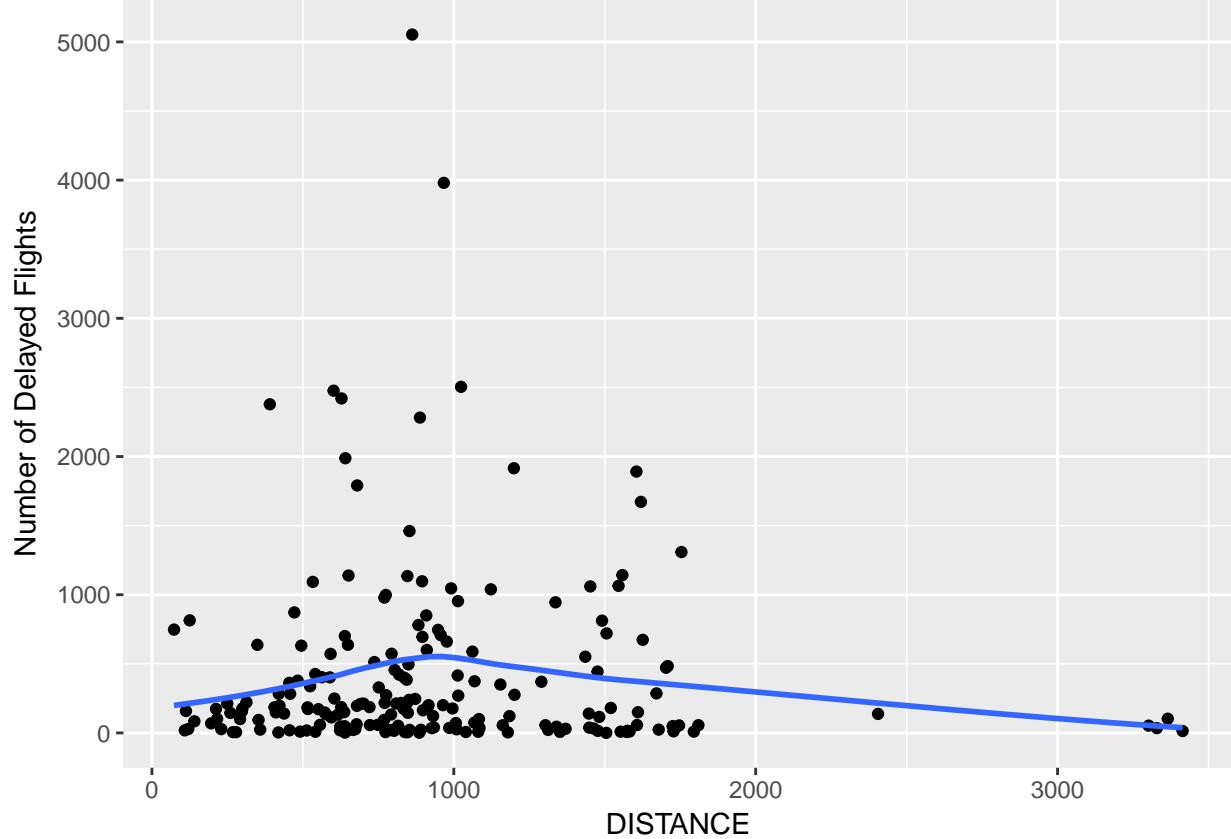
## Conditional probability a flight is late given the destination is DEN: 0.157700255366763
message("conditional probability a flight is late given it is leaving DEN: ", P_late_fromDen)

## conditional probability a flight is late given it is leaving DEN: 0.186637015251233
```

The two plots show, respectively, how frequently flights are delayed a given amount of time when a flight is leaving or landing at the Denver airport. Both graphs have a similar distribution, however, flights into DEN have many more flights with a 20minute delay or greater. Calculated conditional probability demonstrate the probability of late arrival given the destination is Denver is 16%, and the probability of a late departure given a flight is leaving Denver is 19%.

Zhenlong Li

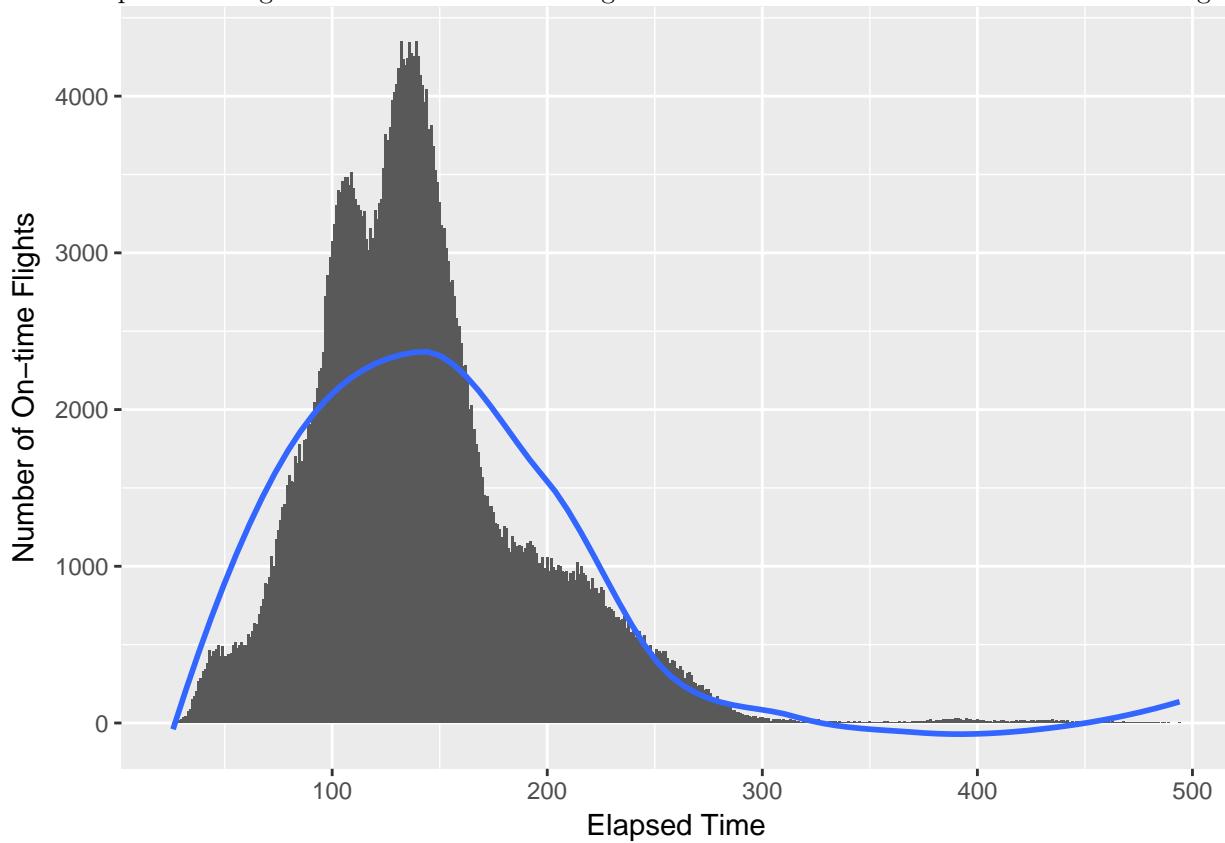
Here's a plot showing that most delayed flights occurred where flying distance is between 750 mile to 1000 miles:



```
##           n  
## 1 0.3515244
```

And the probability of that flying distance between 750 mile to 1000 miles will have 35% to be delayed.

Here's a plot showing that the actual air time greater than 300 mins will have less on-time flights:



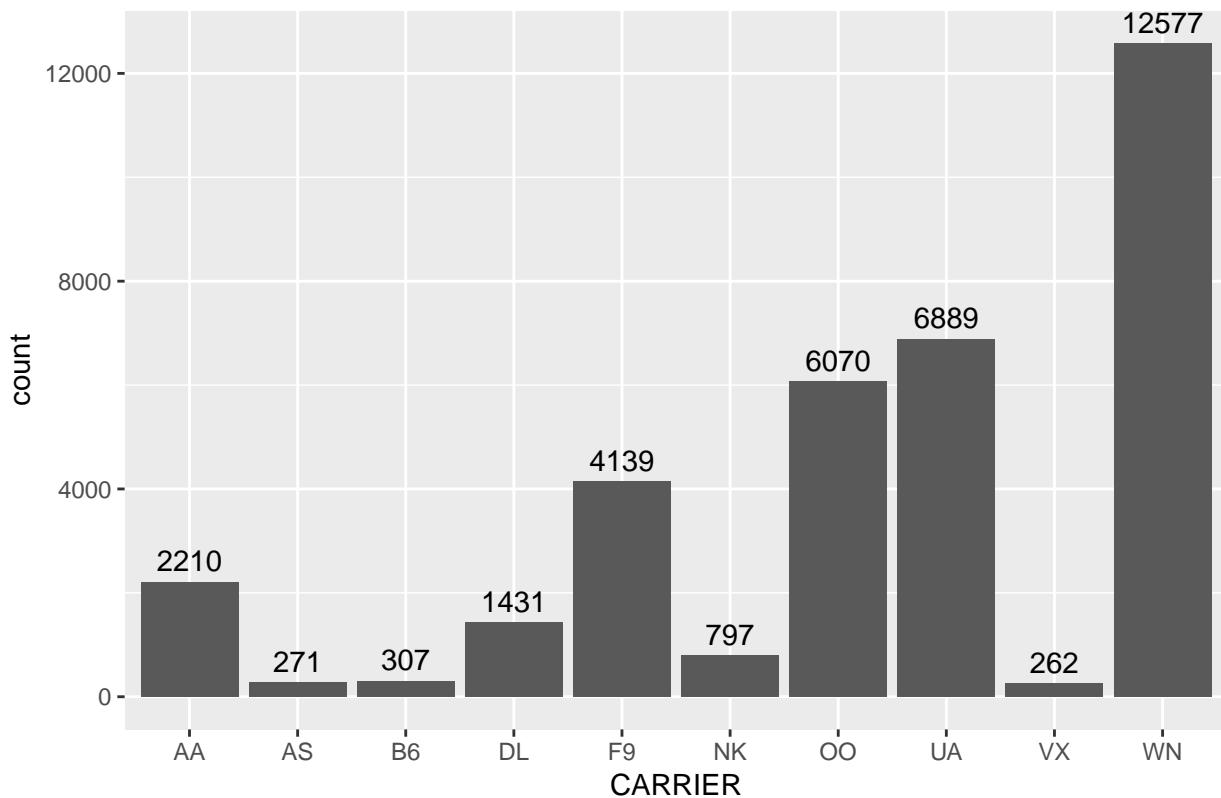
```
##          n  
## 1 0.006547695
```

And the probability of that the actual air time greater than 300 mins will have 0.65% to be on-time.

Scott Baker

First, I take a look at the number of flights that arrived late (more than 15 minutes) into Denver based on airline. Here is the graph:

Number of ARR_DELAY >= 15 flights into Denver per Airline



The first major observation is that airlines “WN”, “UA”, and “OO” have the most flights delayed landing. * Southwest (WN): 12577 * United (UA): 6889 * SkyWest (OO): 6070

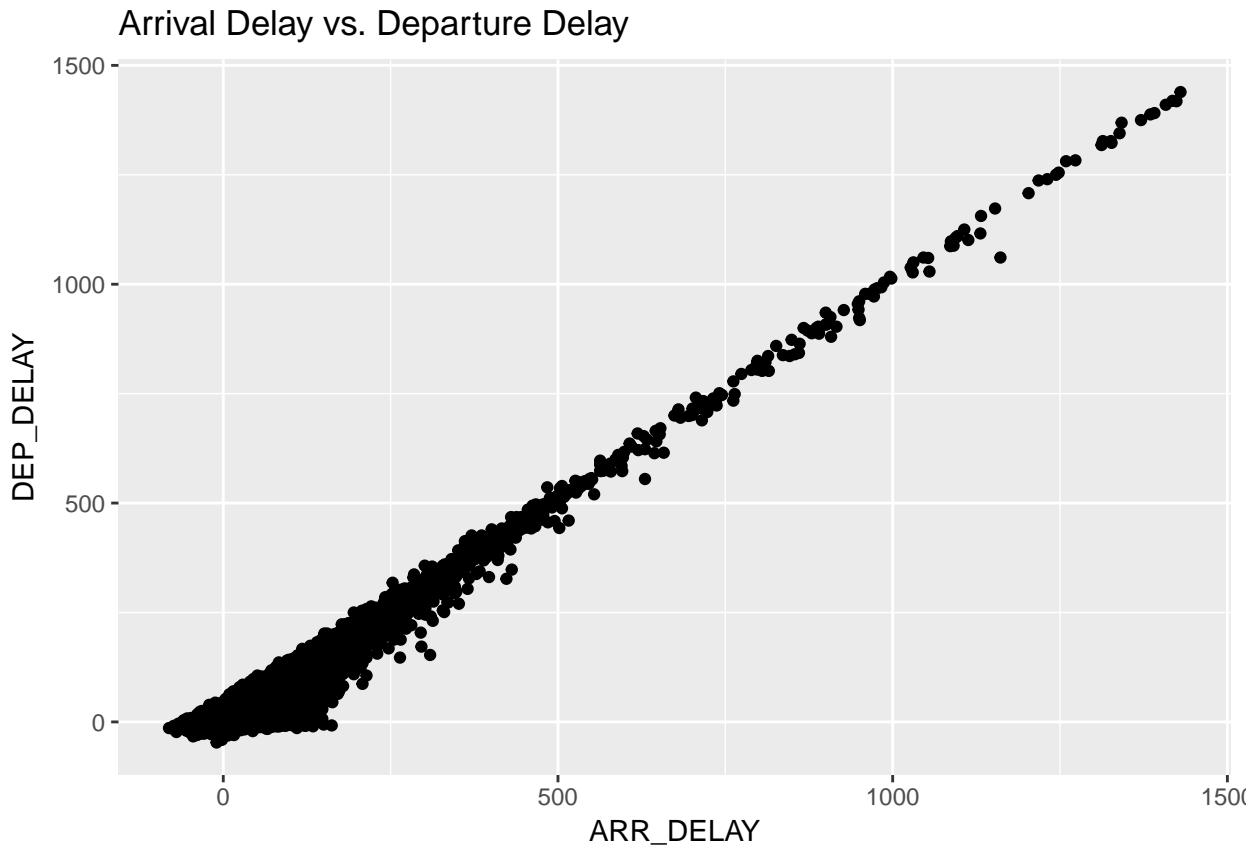
Although this confirms personal experience that some United and Southwest flights I’ve been on have been delayed, this does not provide any insight into the probabilities that a certain airline will have a delayed arrival. Furthermore, let us look at the actual proportion that a certain airline’s flight to Denver is late. Observe the following table:

```
## .
##      AA        AS        B6        DL        F9        NK        OO
## 0.2005445 0.1600709 0.2731317 0.1424164 0.1952543 0.2023870 0.1415644
##      UA        VX        WN
## 0.1113158 0.2536302 0.1838340
```

Here, we can see that “B6”, “VX”, and “NK” have the highest rate of late flights. * JetBlue (B6): 27.31% * Virgin (VX): 25.36% * Spirit (NK): 20.24%

From this, we find that actually JetBlue, Virgin, and Spirit have the most delayed arrivals with respect to total flights-NOT United, Southwest, and SkyWest.

The second part of my analysis is to show that there is a strong correlation between the departure delay and the tardiness of their arrival:



```
## [1] 0.9569532
```

From this we have two results. First, the correlation value between arrival delay (in Denver) and departure delay is .957, which is a strong correlation. Second, the implication of this is that there sometimes might not be anything DIA can do better to ensure more flights are on time—it could all be a product of the departure airport and their inability to get planes off in time. Although, this observation is generic and more specifics would be needed to confirm.

Contributions

- Lindsay: Created density plots to see how flight delays were impacted given that flights were either into or out of Denver. Calculated conditional probabilities that a flight was late given the destination or origin of a flight was DEN.
- Lexie: Created individual plot and decription of graph. Also helped create the conclusion plot with help conclusion paragraph and recommendation ideas.
- Li: Creted two plots to see how distance and actual elapsed time affected whether a flight is delayed or on-time. Checking my findings by using calcluating the probability of each event.
- Scott: Besides the normal organization of the git and knitting/submitting the final file, I took care of the recommendation to DIA officials. I also examined airline-by-airline arrival delay.