Econ294A Final

Kevin Troxell

March 16, 2016

Weather

Weather may be a predictor of flight delay times. The most sensible prediction is that windgust is positively correlated with delay times. I ran a non-parametric regression to see how windgust affects departure time.

```
scatter <- ggplot(wind, aes(wind_gust, dep_delay)) +
  geom_smooth() +
  geom_abline() +
  ggtitle("Windgust and Delay Times") + ylab("Minutes Delayed") + xlab("Wind Gust")
scatter
```

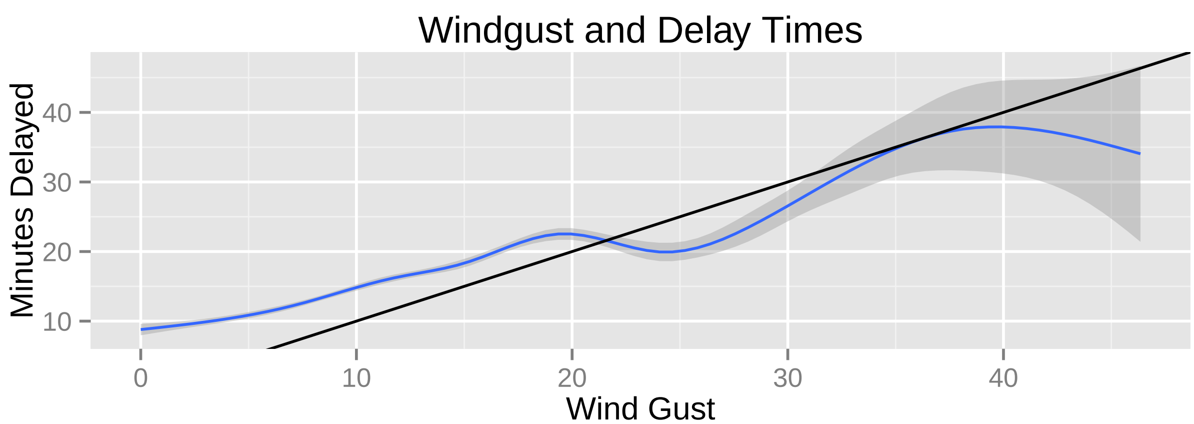

Figure 1:

I also ran a linear regression: $dep\_delay = \beta_0 + \beta_1 wind\_gust + u_i$ to see how windgust affects departure delays linearally.

```
winddelay <- glm(dep_delay ~ wind_gust, wind, family=gaussian)
```

The coefficient on windgust is predicted to be .65, indicating that as windgust speed increases by 1 mile per hour, departure delay increases by .65 minutes. This is statistically significant at the 99% level.

Carrier

I also wanted to see how carriers performed based on their departure delays. For this, I looked flights that were delayed over one hour. The histogram below shows the frequency of flights delayed and the frequence of carriers that were delayed. There doesn't appear to be any carriers that have an overwhelming number of delays.

```
flights_planes <- flights %>% full_join(planes, by="tailnum")
very_delayed <- flights_planes %>%
  filter(dep_delay = dep_delay>=60)
carrierhist <- ggplot(very_delayed, aes(dep_delay, fill=carrier)) +
  geom_histogram(binwidth=20) +
  xlim(50,500) + ggtitle("1hr+ Delayed Flights by Carrier") + xlab("Minutes Delayed")
carrierhist
```

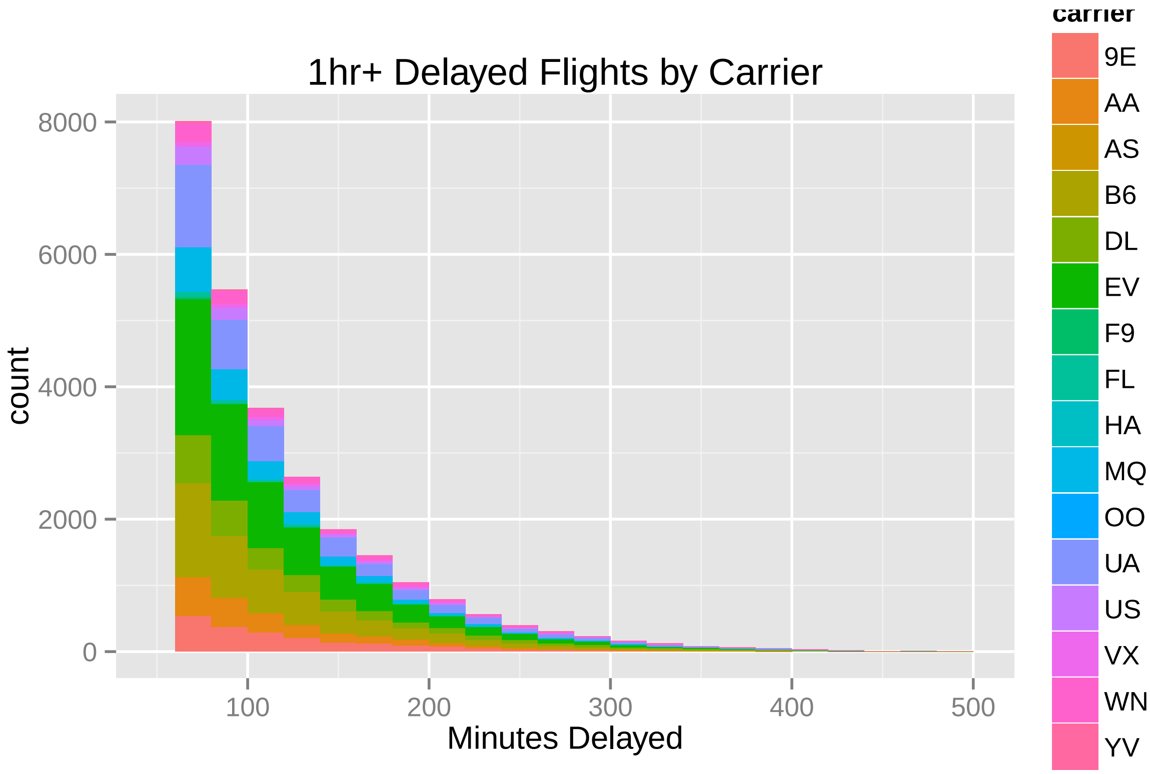

Figure 2:

I ran a linear regression to see how departure delays depend on the carrier. $dep\_delay = \beta_0 + \beta_i carrier_i + u_i$

```
carrierdelay <- lm(dep_delay ~ carrier, flights)
```

Here, we can see that the intercept, carrier 9E, on average departs 16.725 minutes late. The most notable carriers are US which departs on average 3.785 minutes late, and carrier F9 which departs on average 20.2 minutes late.

Time of Day

Lastly, I regressed the hour of the day on departure delays to see if later flights were more likely to be delayed: $dep\_delay = \beta_0 + \beta_1 hour + u_i$

```
delayhour <- lm(dep_delay ~ hour, flights_planes)
```

Here, I found the coefficient on hour to be 2.15 meaning that as the time of the day increases, the departure delay increases. This makes sense because I would expect that delays would build up throughout the day.

```
scatter <- ggplot(flights_planes, aes(dep_delay, hour, color=origin, size=seats)) +
  geom_point(alpha=1/2) + ylab("Departure Delay") + xlab("Hour") +
  ggtitle("Hour of the Day and Departure Delays")
scatter
```

In the scatter on the next page, I measured how hour of the day affects departure delays. It is difficult to tell with a graph, but it seems that as hour increases, delay increases. The airport of departure is denoted by the dot color and the size of the airpline is denoted by the dot size.

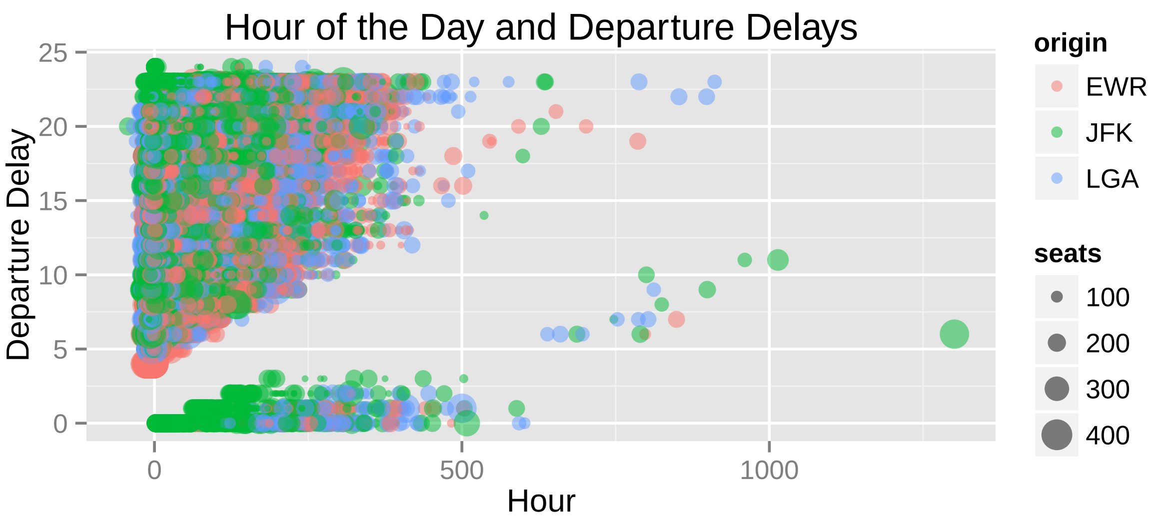
Hour of the Day and Departure Delays
Departure Delay
25
20
15
10
5
0
0
500
1000
Hour
origin
EWR
JFK
LGA
seats
100
200
300
400

Figure 3: