DATA 5300: Analysis of United Airlines Departure Delays

Team Member: Vincent Chan, Muykhim Ing, Jimmy Nam, Alex Song

Introduction

Flight delays are among the most significant factors influencing customer satisfaction in the airline industry. For United Airlines, reducing delays on flights departing from the New York City area is essential not only for improving operational efficiency but also for enhancing the overall customer experience. To explore the causes and patterns of these delays, our team analyzed data from the **nycflights13** dataset, which contains detailed records of all flights departing from New York City in 2013. While the dataset is comprehensive, this project focuses specifically on information relevant to United Airlines.

Our analysis examines how departure delays are related to several key factors, including time of day, time of year, and weather conditions such as temperature, wind speed, precipitation, and visibility. By identifying patterns and relationships across these variables, we aim to provide actionable insights that can help United Airlines minimize delays, optimize flight operations, and improve customer satisfaction.

Data Methodology

The data used in the project was *nycflights13* package in R, which contains multiple datasets, including information on flights, airlines, weather, and airports for the year 2013. To focus on United Airlines, we filtered only flights operated by UA from the *flights* dataset.

We then, merged the flight data with the *weather* dataset using shared keys.

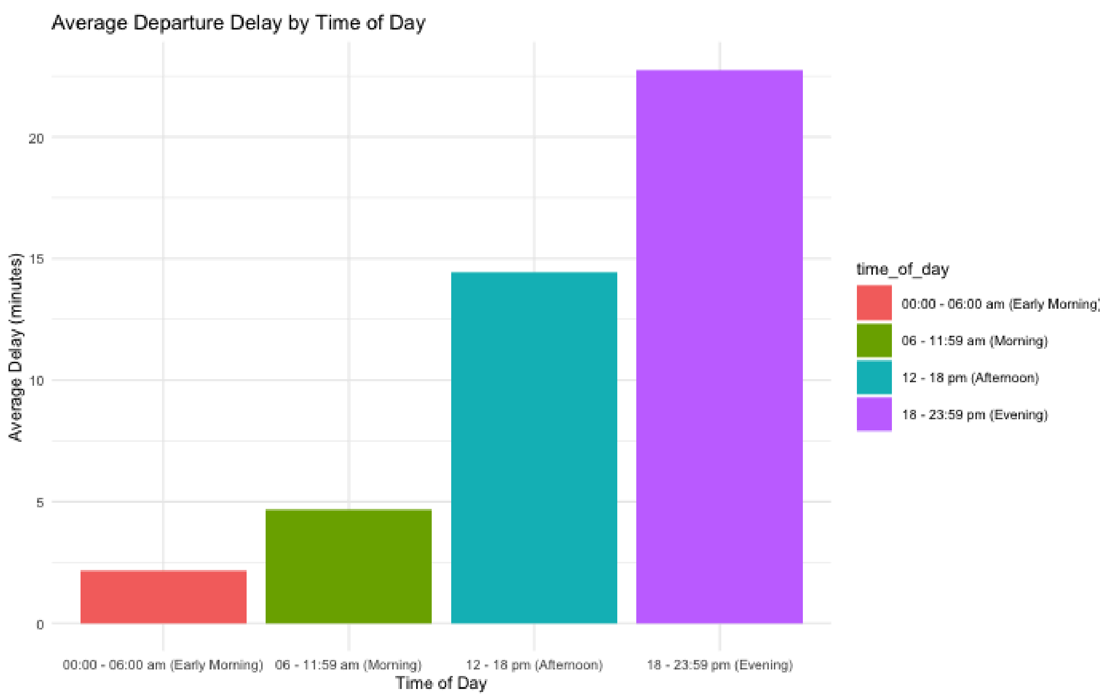
'Year', 'month', 'day', 'hour', 'origin'

This join allowed us to attach weather conditions such as temperature, wind speed, visibility, and precipitation to each flight record. After merging, we cleaned the dataset by removing flights with missing values in variables like *dep_delay* and *visib*. Depending on the comparison, we created new categorical variables to simplify analysis. Finally, we saved the cleaned dataset *United_flights_weather_joined* containing all United Airlines flights departing from NYC in 2013 with complete weather data.

```
United_flights_weather_joined <- United_flights %>%
  inner_join(weather, by = c("year", "month", "day", "hour", "origin")) %>%
  drop_na(dep_delay, visib)
glimpse(United_flights_weather_joined)
```

Results

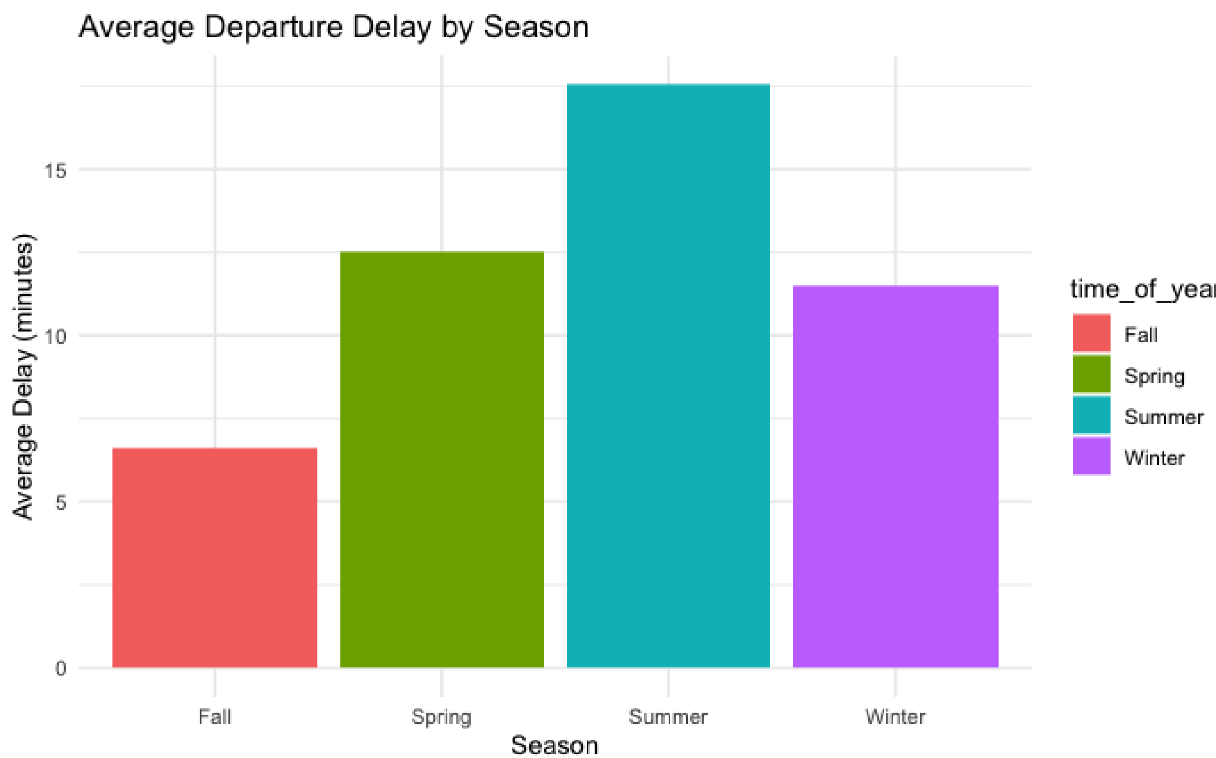
1. **Time of Day**



The Bar plot of Departure Delays by Time of day shows a clear trend. Flights in the Early Morning (00:00 - 06:00am) experience the shortest delays, indicating they generally depart on time. Delays increase slightly for Morning flights (6:00 - 11:59am) and become more noticeable in the Afternoon (12 - 18pm). Evening flights (18 - 11:59pm) have the longest delays, suggesting that delays tend to accumulate throughout the day, likely due to earlier disruptions affecting later departures.

To assess whether this difference is statistically meaningful, a permutation test was conducted comparing Morning and Evening flights using the proportion of delayed flights. The resulting **P-value was 0.00019**, which is extremely small. This indicates that the higher frequency of delays in the Evening is unlikely to be due to random chance. These findings support the observation from the bar plot that Evening flights are significantly more likely to be delayed than Morning flights.

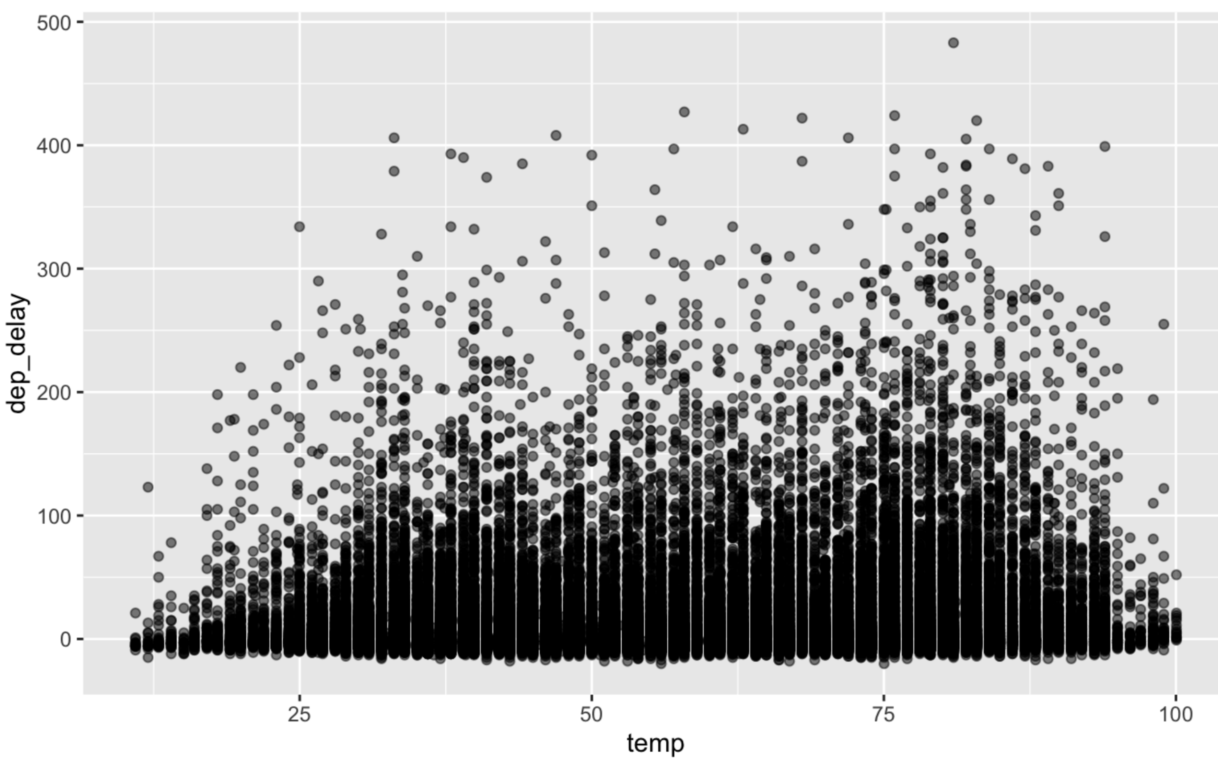
2. Time of Year

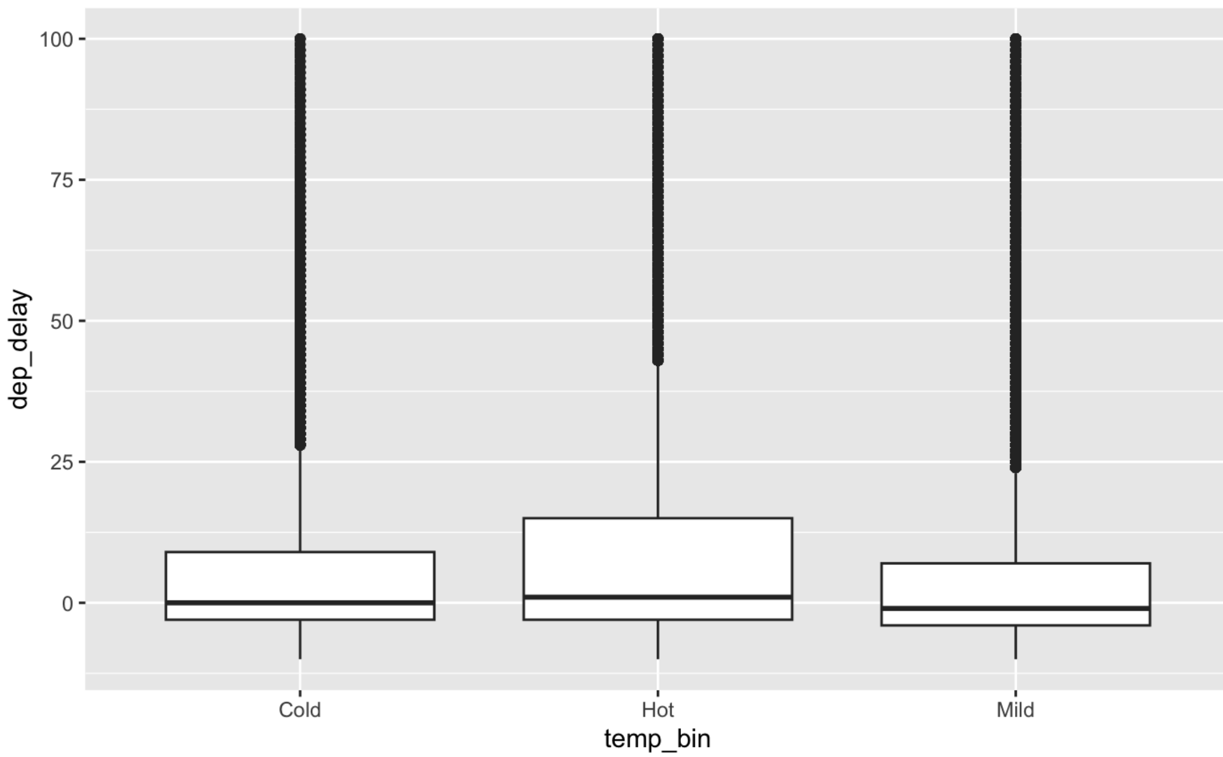


The Bar plot of Average Departure Delays by Season shows that Summer flights experience the longest delays. Spring and Winter have moderate delays, while Fall flights have the shortest delays time. This indicates that seasonal factors, such as weather conditions and increased travel volume, influence flight punctuality, making Summer the most delay-prone season.

A permutation test comparing Summer and Fall flights using the proportion of delayed flights produced a **p-value of 0.00019**. This extremely small value indicates that the observed difference is highly unlikely to have occurred by chance. Therefore, the results provide strong evidence that season has a significant impact on departure delays, with Summer showing the most delays and Fall the least.

3. Temperature:





When looking at how temperature affected departure delays, most UA flights left close to their scheduled time regardless of the weather. The scatter plot showed no strong pattern overall, but there were more long delays during very hot days.

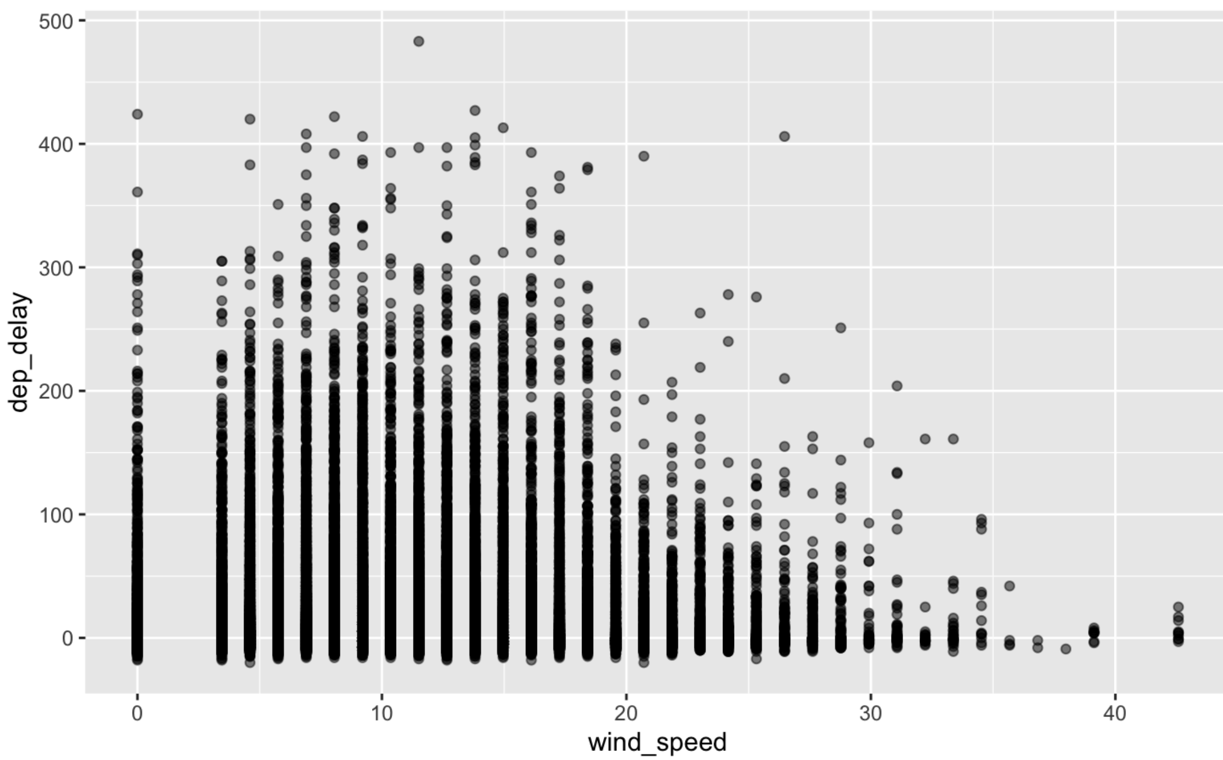
After grouping temperatures into three categories - Cold (below 40°F), Mild (40–75°F), and Hot (above 75°F), the boxplot showed that flights during hot and cold weather had slightly higher median delays and a wider range of delay times compared to mild weather. The histogram partially supported this trend. Most flights still departed near schedule, but there were noticeably more extreme delays in hot conditions.

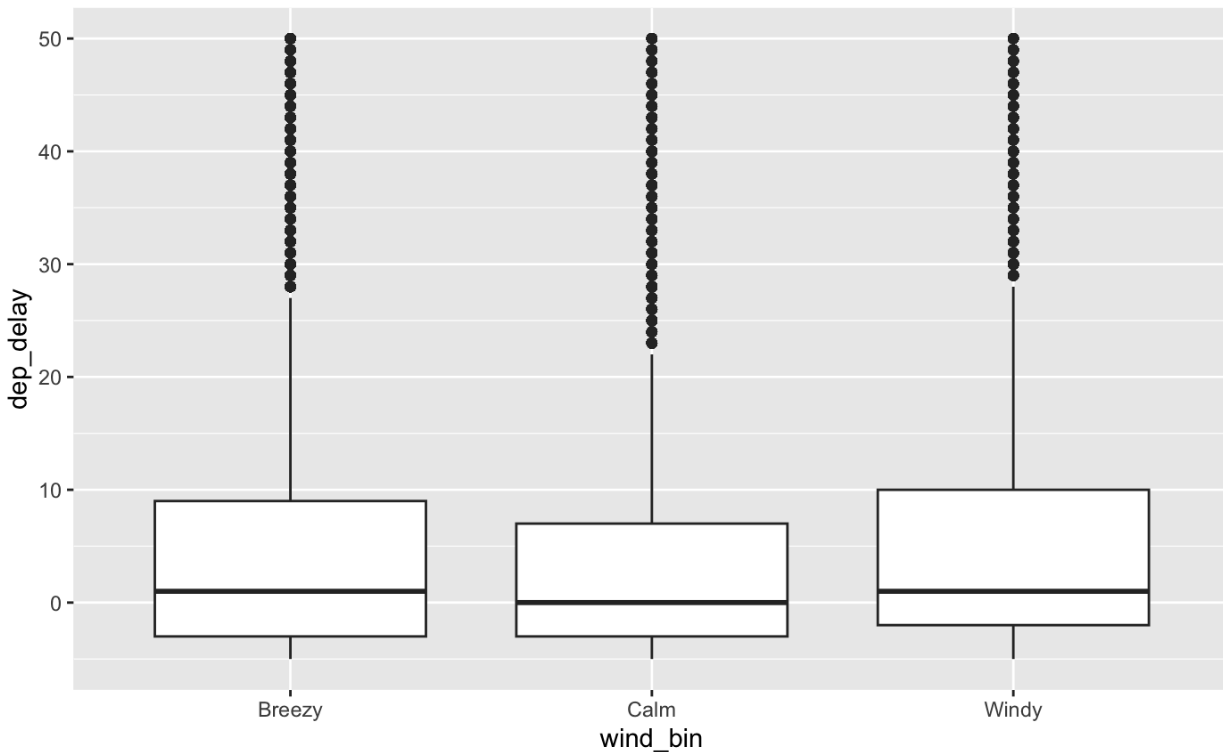
To see if these differences were meaningful, permutation tests were used. The comparison between hot and mild days gave a p-value of 0.0002, which shows that the longer delays on hot days were statistically significant. In contrast, the comparison between cold and mild days gave a p-value of 0.56, meaning the difference was not statistically significant.

The results show that extreme heat is linked to longer delays. This likely happens because hot air is less dense, which reduces lift and engine thrust. Under these conditions, an aircraft may not have enough performance for the available runway length or to safely clear nearby obstacles.

Extremely cold weather, on the other hand, did not appear to have a strong effect on departure times. One possible reason is that when temperatures are very low and snow is present, flights are more likely to be canceled instead of delayed.

4. Wind speed:





The scatter plot between departure delays and wind speed showed that flights generally had small delays across all wind speeds, with no strong trend between the two variables. When flights were grouped into three categories - Calm (below 8 mph), Breezy (8 to 15 mph), and Windy (above 15 mph), the boxplot showed very similar median delays in all three groups. Delays became a little more variable during windy conditions, but the overall pattern stayed about the same.

A permutation test comparing windy and calm conditions gave a p-value of 0.0002, suggesting that the difference between these groups was statistically significant. While the difference was not large, the result indicates that stronger winds can increase departure delays. Strong winds can make it difficult and dangerous for pilots to control the aircraft during takeoff, which can lead to longer delays or temporary pauses in departures until conditions improve.

5. **Precipitation**

Across all United Airlines flights departing from New York City in 2013, flight delays were fairly common even under normal weather conditions:

- About one in three flights (33.8%) departed slightly late (up to 30 minutes behind schedule).
- About one in eight flights (13.1%) departed very late (more than 30 minutes behind schedule).

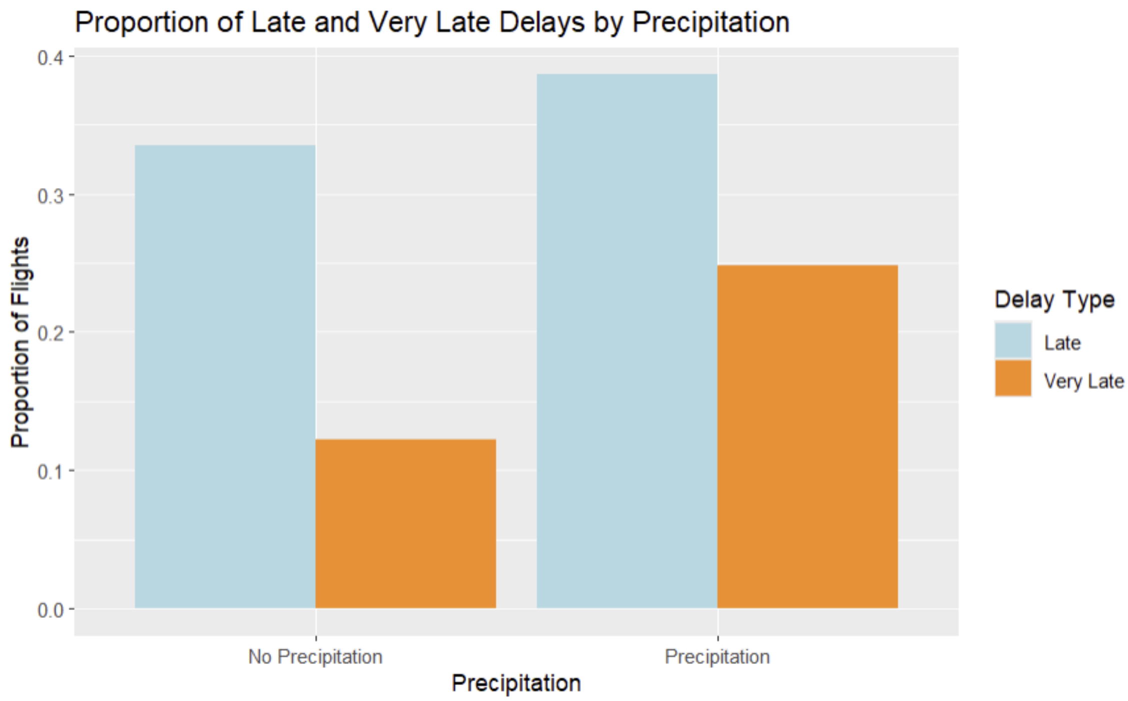
To explore how the weather affected these delays, flights were compared based on whether they experienced precipitation.

Precipitation	**Late (Delay < 30 mins)**	**Very Late (Delay > 30 mins)**	**Number of Flights**
Overall	0.338	0.131	57,686
No Precipitation	0.335	0.122	53,645
Precipitation	0.385	0.248	4,041

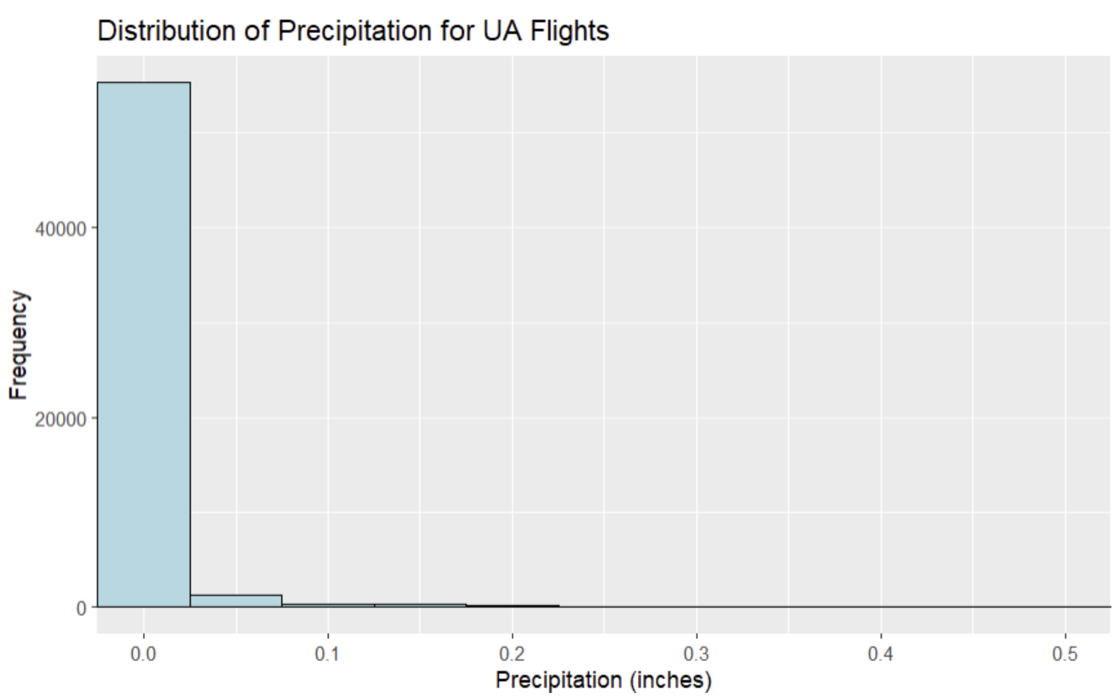
The results show a clear difference. Flights that encountered precipitation had higher delay rates across the board:

- **38.5%** departed late, compared to **33.5%** without precipitation.
- **24.8%** departed very late, compared to just **12.2%** under dry conditions.

In other words, the proportion of very late departures **more than doubled** on rainy days, underscoring the strong effect of wet weather on flight timeliness.



A grouped bar chart visualized this pattern. The x-axis represented precipitation conditions ("No Precipitation" vs. "Precipitation"), while the y-axis showed the proportion of flights by delay type. The chart clearly indicated that both **Late (light blue)** and **Very Late (orange)** delays were more common during precipitation, confirming the greater likelihood of delays in rainy weather.



To understand how often precipitation occurred, a histogram of precipitation levels was also created. It showed that most flights took place in little to no precipitation, meaning that rainy conditions were relatively rare. However, despite the smaller number of rainy flights, their much higher delay rates make the relationship between precipitation and delays particularly noteworthy. In this histogram, the x-axis represented precipitation (in inches), and the y-axis showed the number of flights.

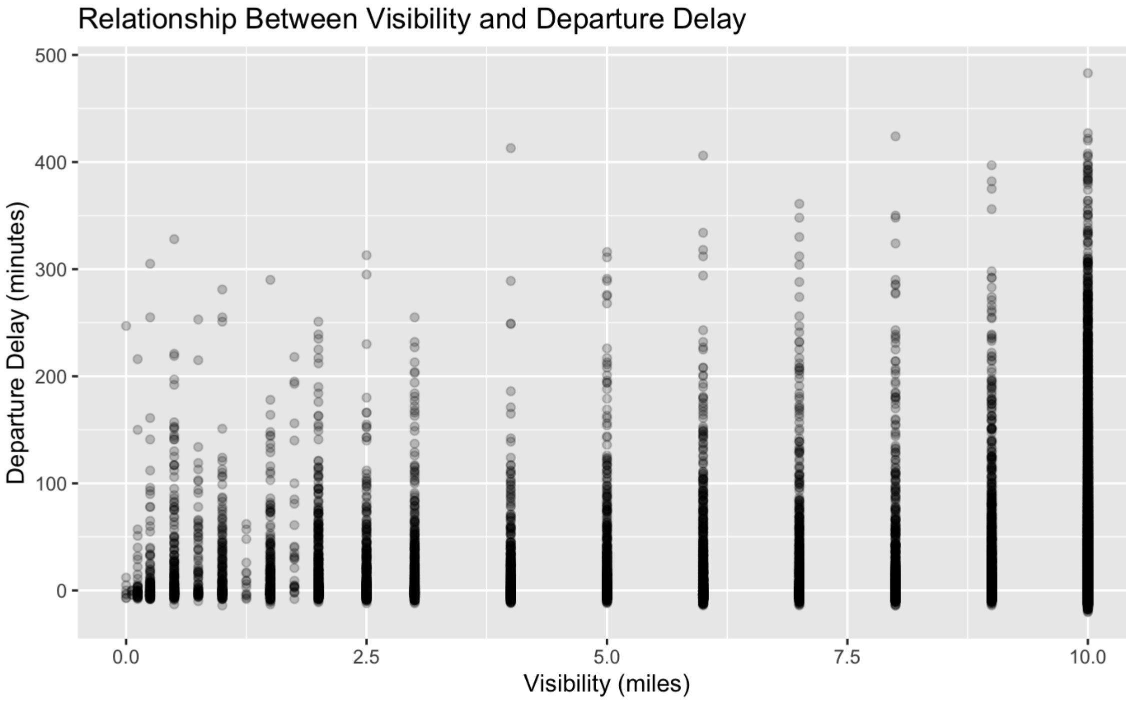
To test whether these differences were statistically significant, a permutation test was performed comparing the overall proportion of delayed flights (late or very late) between precipitation conditions.

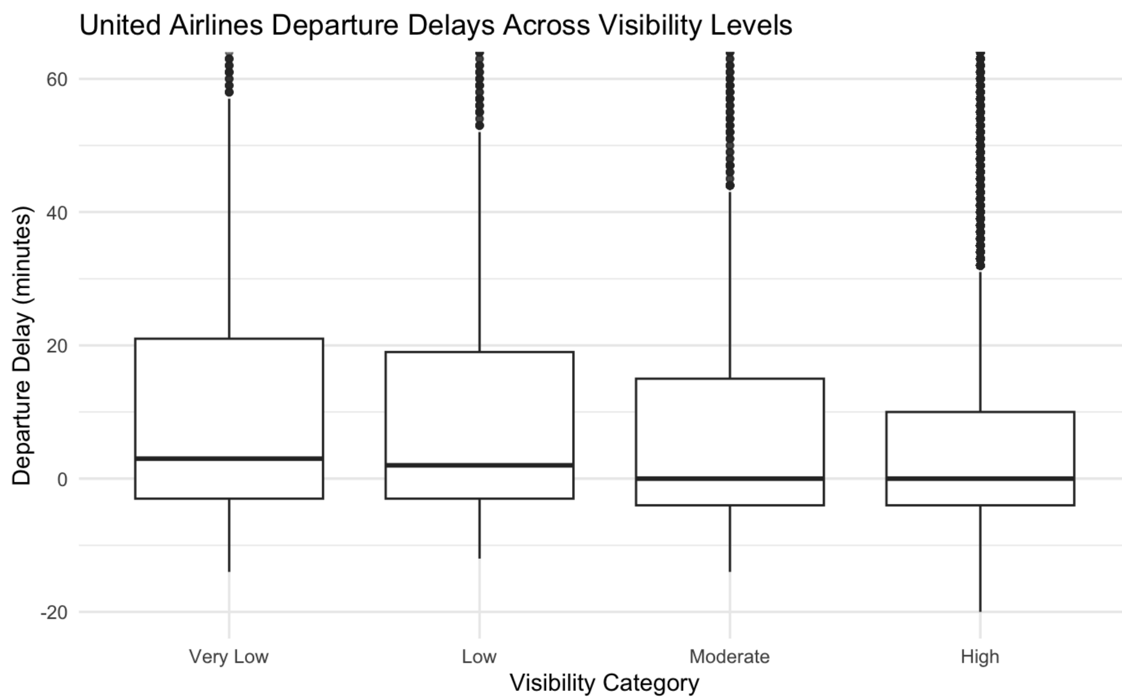
- **Null Hypothesis (H_0)**: There is no difference in the proportion of delayed flights between precipitation and no-precipitation conditions.
- **Alternative Hypothesis (H_1)**: There is a difference in the proportion of delayed flights between the two conditions.

The observed difference was 0.177, meaning flights with precipitation were **17.7%** more likely to experience any delay. Even though only 4,041 flights occurred in precipitation (compared to 53,645 without), the difference was large and consistent. Running 10,000 permutations produced a **p-value of 0**, providing strong evidence against the null hypothesis. This result confirms a statistically significant difference in departure delays between rainy and dry conditions.

In summary, precipitation had a clear and measurable impact on flight timeliness. United Airlines flights departing in rainy weather were significantly more likely to be delayed, and those delays tended to be longer. This finding aligns with the general prediction that wet weather can disrupt airport operations, from ground handling and taxiing to takeoff clearance, ultimately slowing down flight departures.

6. **Visibility**



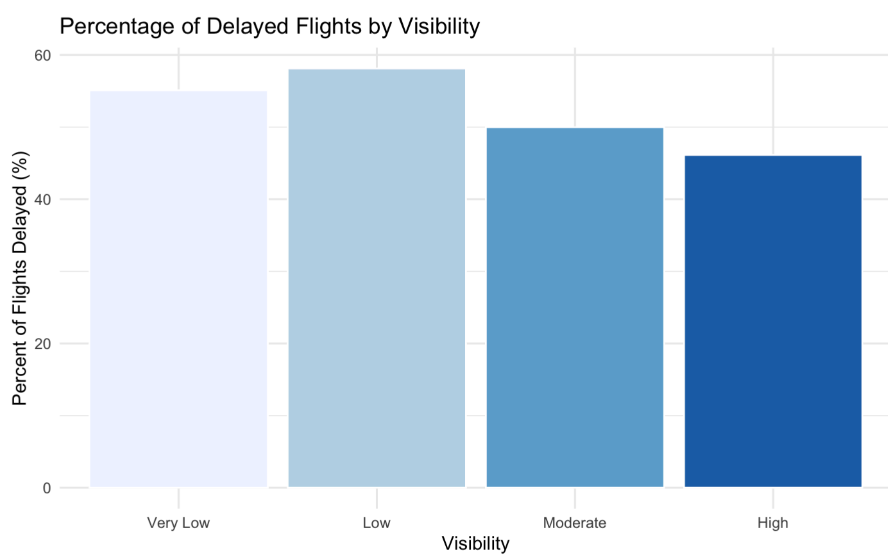


The scatter plot between departure delays and visibility shows that flights tended to have longer delays when visibility was low, especially below around 3 miles. There was still a lot of variation at all visibility levels, but the cluster of higher delays was more noticeable under poor visibility conditions.

Visibilities were divided into four categories:
Very Low: < 2 miles
Low: 2 ~ 5 miles
Moderate: 5 ~ 8 miles
High: > 8 miles

The boxplot showed that delays were generally higher and more spread out during very low and low visibility. Flights under clear, high-visibility conditions had lower median delays and less variation overall.

A permutation test comparing low and high visibility conditions gave a p-value below 0.05, suggesting that the difference between these groups was statistically significant. While the difference wasn't huge, it still suggests that reduced visibility can contribute to longer departure delays. Low-visibility conditions like fog or haze can slow down ground operations and reduce takeoff efficiency, leading to temporary pauses or extended delays until conditions improve.



Discussion

This study aimed to understand what factors most influence United Airlines departure delays using the *nycflights13* dataset. Overall, our analysis found that both time-based and weather-related factors play significant roles. Delays were shortest early in the morning and increased throughout the day, suggesting that small disruptions accumulate as the schedule progresses. During the summer, flights experienced the most delays, likely because of increased passenger traffic and more frequent storm activity.

Weather conditions also showed clear effects. Precipitation and low visibility had the strongest relationships with longer and more variable delays. Temperature and wind speed had smaller but still noticeable impacts, especially under extreme conditions. These results align with real-world operations, where reduced visibility, rain, or wind can slow down ground procedures, limit runway capacity, and lead to temporary pauses in departures.

While the dataset provided valuable insights, there were some limitations. The data was limited to 2013 and did not include factors such as air traffic control delays, maintenance issues, or staffing constraints, which can also affect flight delays. Future analysis could expand by including more recent data, comparing multiple airlines, or using predictive modeling techniques to forecast delays in real time.

In a broader sense, this study highlights how data analytics can support decision-making in the airline industry. Understanding when and why delays happen allows airlines like United to plan more efficiently, allocate resources better, and improve customer experience. Continued work in this area could help reduce operational costs, increase on-time performance, and make air travel more reliable overall.