DECEMBER 1, 2025

GAIN PER FLIGHT

Analysis Report

APPLIED STATISTICAL INFERENCE & EXPERIMENTAL DESIGN (DATA 5300)

INTRODUCTION

Delays in commercial air travel are a widespread phenomenon experienced across the global aviation industry. Most time, a delayed departure does not necessarily mean that the flight will have a delayed arrival – it is often possible for airlines to recover time when they are airborne by making changes to the routes, adjusting speed, or some other modifications in air traffic conditions. This project focuses on understanding when and how much time can be made after a delayed departure.

In this report, we will be using the United Airlines flights departing from New York City in 2013 provided as a stand-alone package from the nycflights13 package. The main variable of interest in our analysis is **net gain** which is a generated variable by subtracting arrival delay from departure delay. We also analyze gain per hour, which is a better estimate of longer flight recovery times.

Problem Statement

A major player in the aviation industry, United Airlines, wants to better understand how operational factors affect schedule recovery during flight. Specifically, United Airlines are interested in answering the following questions:

1. Does departing late impact a flight's ability to make up time while airborne?
2. Which top five (5) destination airports show the largest or smallest time recoveries?
3. Do longer flights gain more time, especially when accounting for flight duration?
4. At what level of departure delay – like exceeding 30 minutes – does the ability to make up time in the air begin to noticeably decline?

DATA & METHODOLOGY

Data Sources

Our analysis uses the nycflights13 data, which contains 336,776 flights departing from the three (3) different airports in New York City. Given the scope of our analysis, we focused on only flights that are operated by United Airlines (UA) with prescribed carrier code as “UA”.

Data Preparation and Cleaning

To get the dataset ready for analysis, we first brought the flight information into our workspace and checked it for missing information. Some of the flights did not have recorded departure or arrival times, which usually implies that they were cancelled. Since cancelled flights do not provide any insight into time gained in the air, they were removed from our analysis. Additionally, because calculating gain per hour requires a valid airtime value, flights with missing or zero airtime were excluded from the dataset.

Next, a measure called net gain was created. The measure indicates how much airtime a flight made up while travelling. The net gain was calculated by looking at how late a flight took off compared to how late it arrived. For example, if a plane left late but arrived less late, it shows that the flight gained time while in the air.

Since we also wanted to analyze how efficiently flights recovered time, we calculated gain per hour by dividing the net gain by flight airtime (converted to hours).

An indicator was also added to the table to identify whether a flight departed late (delay greater than 0 minutes) or departed very late (more than 30 minutes). Also, to compare performance across different flight lengths, the flights were separated into two categories – short and long categories. We utilized the median airtime as the dividing point because the median is less affected by unusually long or short flights, making it a more reliable threshold than the average.

Given that we also needed to report the five most common destination airports for United Airlines flights from New York City, we wanted to show full airport names instead of just abbreviations. To do this, we combined our flight data with a separate airport dataset included in the **nycflights13** package, which provides the full names for each airport code.

After the merge, we filtered the dataset to keep the columns(variables) relevant to our analysis, ensuring the final data was clean, consistent, and focused on our research goals.

Statistical Methods

To better understand the patterns in flight delays and time recovery, we used simple visual tools such as bar charts and boxplots. These helped us see how often certain destinations were flown to and how much time flights gained or lost in the air. We also looked at basic summaries of the data, like average times gained and how much those times varied from flight to flight.

To compare different groups of flights—such as those that departed late versus on time—we used a standard statistical method that tests whether the differences we see are likely real or just due to chance. We looked for results that were **statistically significant**, which means there is strong evidence that the difference is meaningful. In this report, we consider a result significant if there is less than a 5% chance that it happened randomly.

RESULTS

Hypothesis testing was conducted to understand whether the differences we observed between groups were meaningful, or simply due to random chances. We compared groups such as:

- Flights that departed late versus those that departed on-time (not late)
- Flights more than 30 minutes late (very late) versus those that did not depart 30 minutes late.
- The most common destination airport from NYC.
- Shorter flights versus longer flights.

For each of the comparison, we looked at two main measures:

1. **Total time gained:** How many minutes a flight made-up in the air.
2. **Time gained per hour:** How efficiently time was recovered based on flight duration.

To perform the statistical tests, we made use of a Two sample t-tests with a confidence level of 95%, meaning we only consider a result meaningful if there is less than a 5% chance it occurred randomly. Alongside these tests, we also made use of graphs and summary tables to show how flight performance differs across various groups.

Problem 1, Part 1

Domain Problem: On average, do flights that departed late gain more time than flights that departed on-time (not late)?

Result Interpretation:

The two sample t-test results show:

- Test statistic t = -10.749
- p-value < 2.2e-16
- The 95% confidence interval: [-2.040805, -1.411308]

The confidence interval indicates that flights that depart late gain about 1.4 to 2.0 minutes *less* airtime on average compared to flights that depart on time. Because the entire interval is negative and the p-value is extremely small, we reject the null hypothesis of no difference in mean time gained.

This means the difference is statistically significant: late-departing flights make up less time in the air, not more.

Problem 1, Part 2

Domain Problem: On average, do flights that depart very late (past 30 minutes) gain more time than those that did not?

Result Interpretation:

The two sample t-test results show that:

- Test statistic t = -6.2953
- p-value = 3.215e-10
- 95% Confidence Interval: [-2.21512, -1.268195]

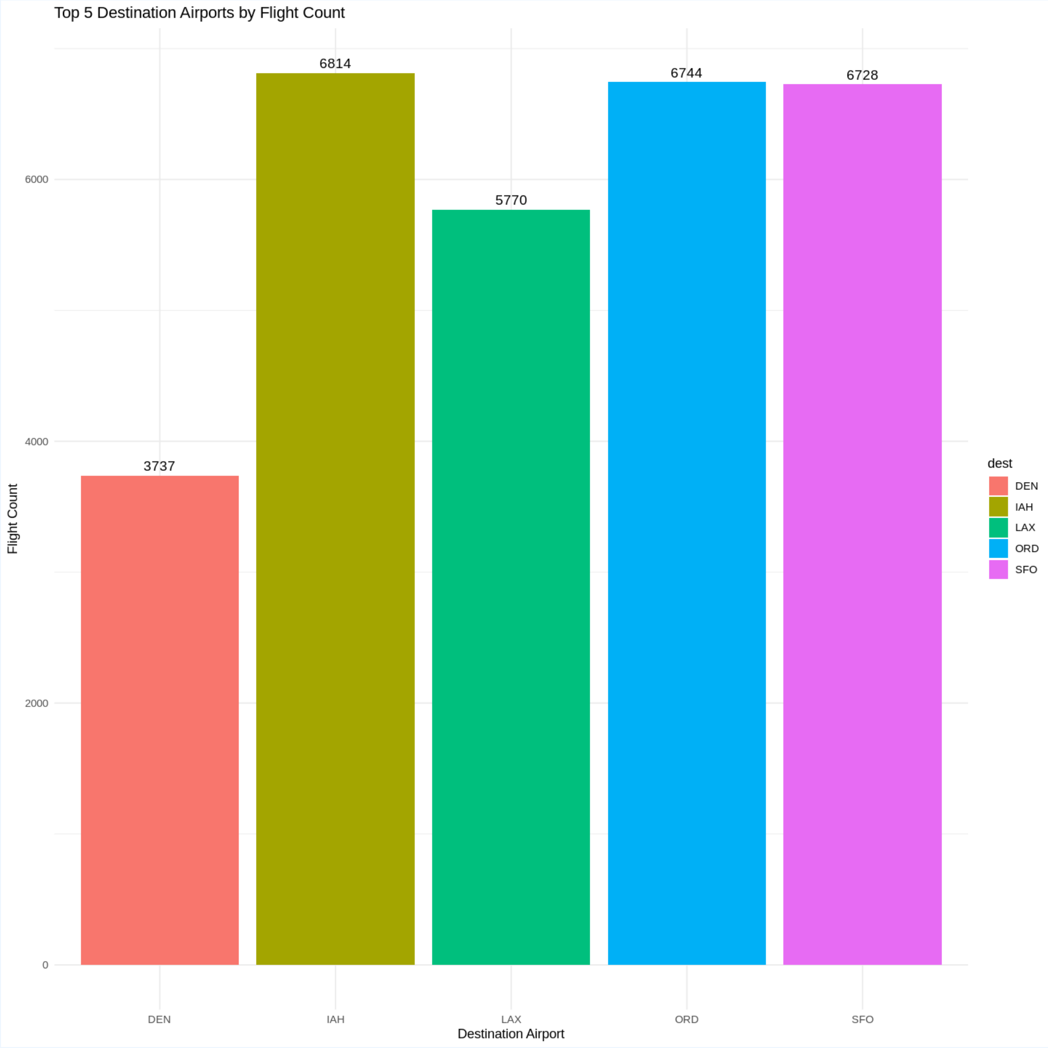
The confidence interval shows that very-late departures gain about 1.27 to 2.22 minutes less airtime on average than flights that do not depart very late. Again, the negative interval and small p-value mean we reject the null hypothesis of no difference.

This indicates that flights departing more than 30 minutes late actually make up 1.2–2.2 minutes less in the air compared to those leaving earlier.

Problem 2, Part 1

Domain Problem: What are the five most common destination airports for United Airlines flights from New York City?

Result Interpretation:



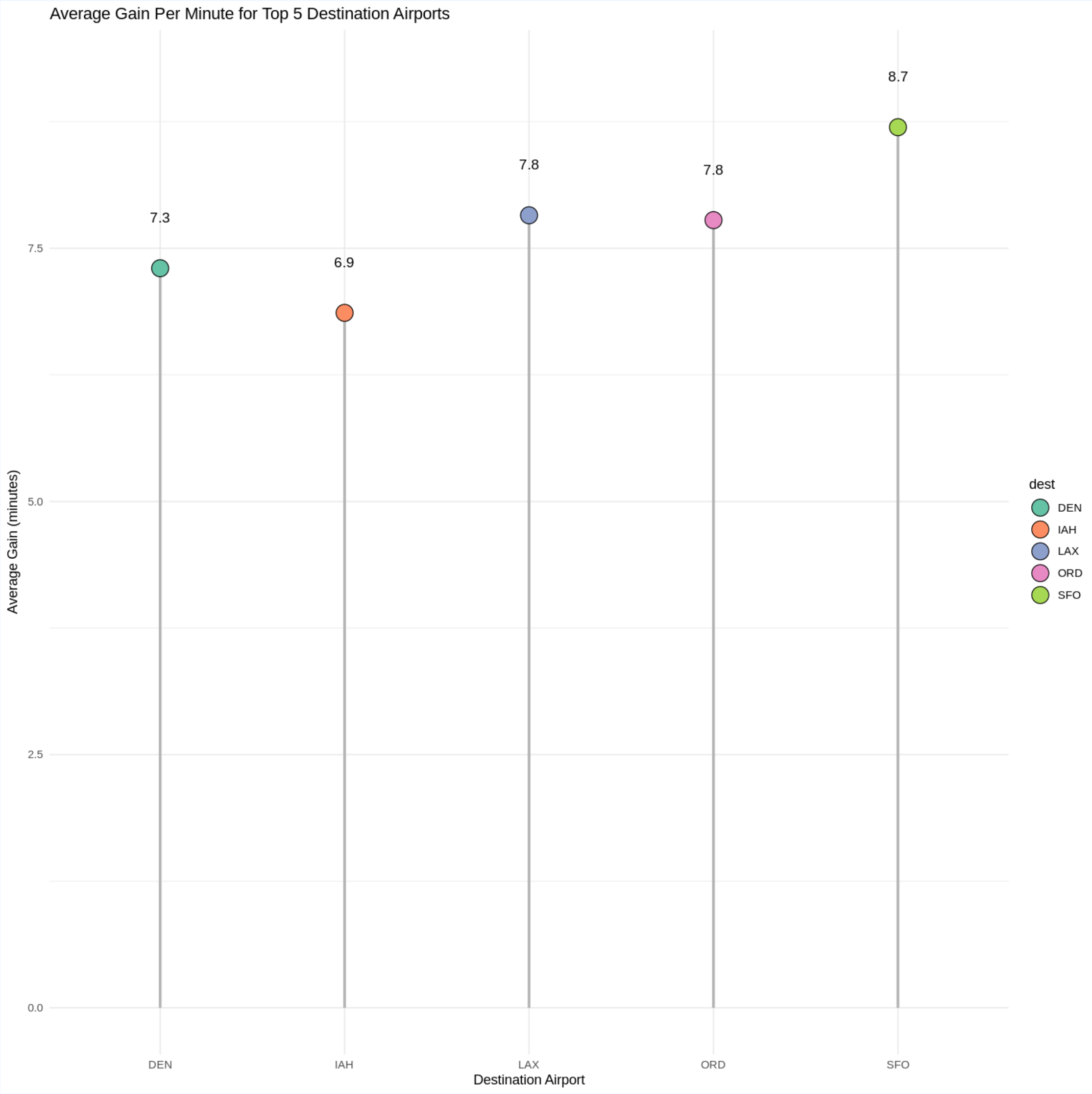
The table shows that the five most common destinations for United Airlines flights departing from New York City are Houston, Chicago, San Francisco, Los Angeles, and Denver. Houston has the highest number of flights at 6,814, followed closely by Chicago with 6,744 and San Francisco with 6,728. Los Angeles and Denver have slightly fewer flights, at 5,770 and 3,737 respectively. Overall, most flights are concentrated to major airport hubs across the United States, with the top three destinations seeing very similar flight volumes.

Problem 2, Part 2

Domain Problem: Describe the distribution and the average gain for each of these five airports.

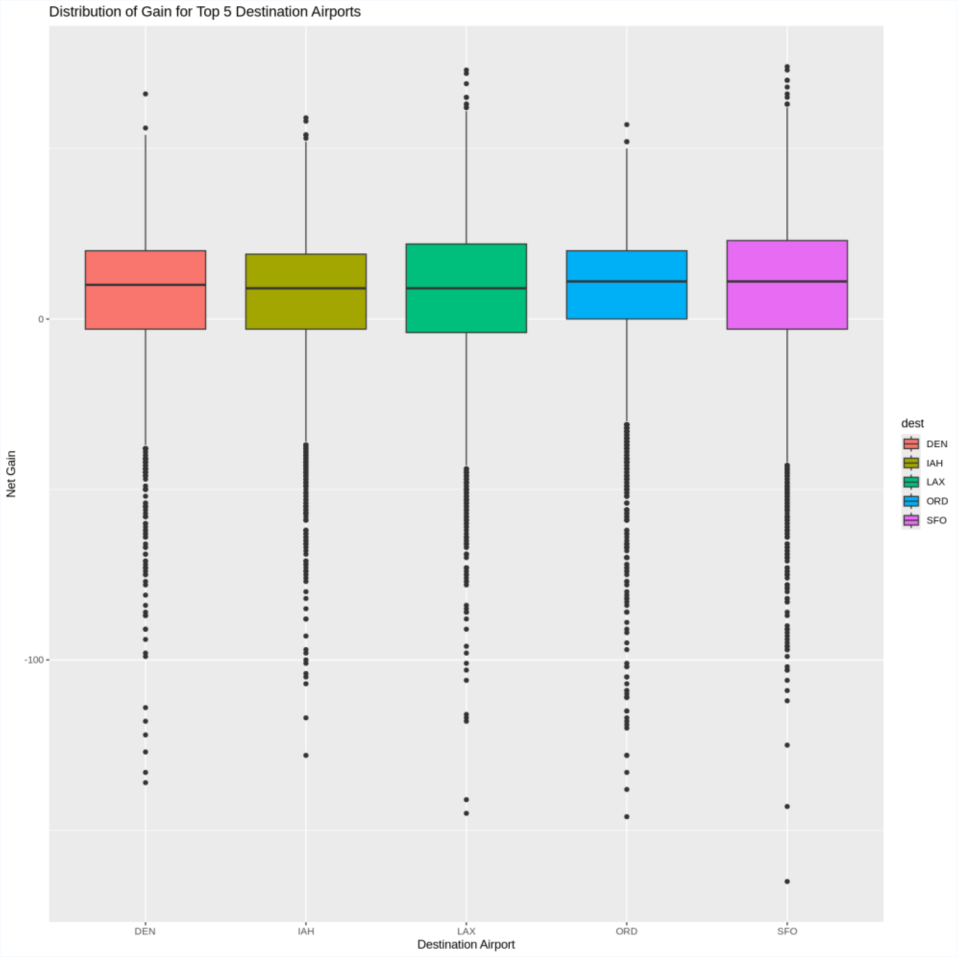
Result Interpretation:

Average Gain for Top 5 Destination Airports:



The histogram shows that flights from New York City to all five destinations generally make up time while in the air. San Francisco stands out with the highest average time recovery of about 8.7 minutes, making it the most efficient route at compensating for delays on the ground. Chicago O'Hare and Los Angeles both recover an average of 7.8 minutes, followed by Denver at 7.3 minutes. Houston has the lowest average recovery among the top five destinations, at around 6.9 minutes.

Gain for Top 5 Destination Airports:



The box plot shows that most flights to these five airports arrive close to their scheduled time, with only small gains or losses in the air. While a few flights arrive much earlier than expected and some are delayed, the majority stay near the planned schedule. Overall, flight patterns are similar across all five destinations.

Distribution and the average gain for each of these five airports:

IAH (Houston George Bush Intercontinental):
There are 6,814 flights to Houston, with an average time recovery of about 6.9 minutes, the lowest among the top five destinations. Most flights arrive close to schedule, though a small number lose a bit of time in the air.

ORD (Chicago O'Hare International):
Chicago receives 6,744 flights, with an average gain of 7.8 minutes. Most flights stay near their scheduled time, with occasional flights recovering a lot of time or experiencing delays.

SFO (San Francisco International):
San Francisco has 6,728 flights and shows the highest average recovery at around 8.7 minutes. Most flights make small gains, and a few recover significant time, making this route the most efficient at compensating for delays.

LAX (Los Angeles International):
Los Angeles sees 5,770 flights, with an average gain of 7.8 minutes. Flight performance is similar to Chicago and San Francisco, with most flights making up time but some experiencing either delays or very large recoveries.

DEN (Denver International):
Denver receives 3,737 flights, with an average gain of 7.3 minutes. Flight times vary more than Houston, showing a mix of flights that recover substantial time and others that encounter delays, typical of longer flights.

Problem 3, Part 1

Domain Problem: Does the average gain per hour differ for flights that departed late versus those that did not?

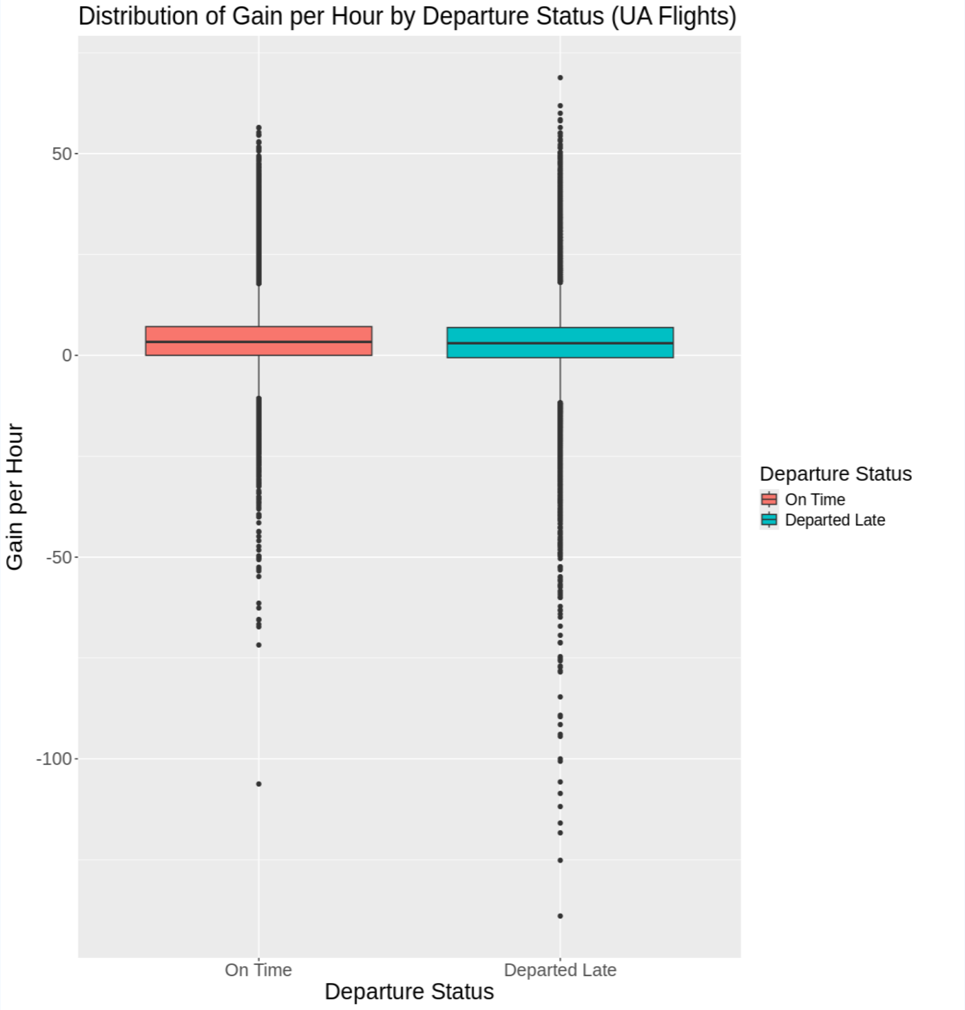
Result Interpretation:

- t-statistic = -11.285
- p-value < 2.2e-16
- 95% Confidence Interval = [-0.947, -0.66]

The two-sample t-tests performed show that:

- Flights that departed late gained about 3.18 minutes per hour on average.
- Flights that departed on-time gained about 3.99 minutes per hour on average.

Distribution of Gain per Hour for Late vs On Time (not late) Flights



The estimated difference ranges from 0.67 to 0.95 fewer minutes per hour for late departures. The associated p-value is extremely small, indicating a statistically significant difference between the two groups.

Problem 3, Part 2

Domain Problem: Does the average gain per hour differ for flights that departed late versus those that departed very late (more than 30 minutes)?

Result Interpretation:

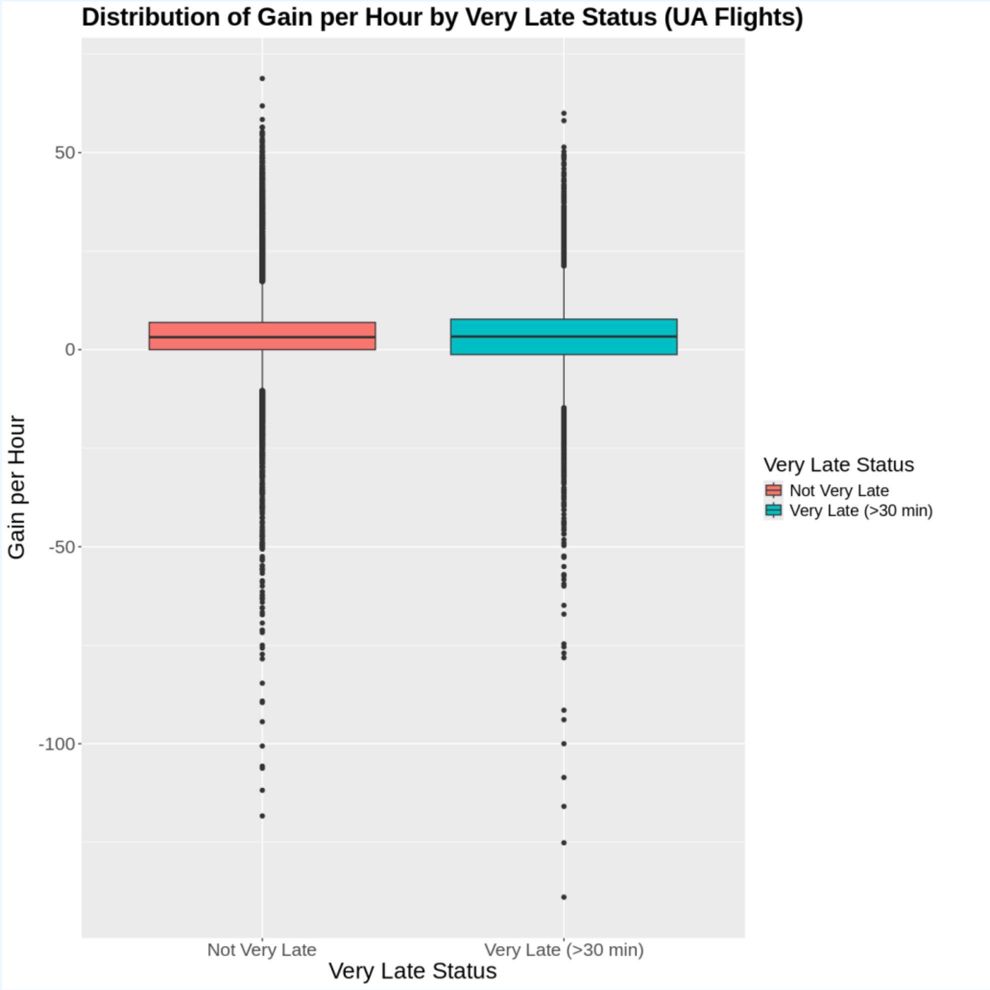
- t-statistic = -4.8323
- p-value < 1.372e-06
- 95% Confidence Interval = [-0.886, -0.375]

The two-sample t-tests performed show that:

- Flights that departed more than 30 minutes late recorded an average gain of 3.06 minutes per hour.

- Flights that did not depart very late recorded an average gain of 3.69 minutes per hour.

Distribution of Gain per Hour for Very-Late vs Not Very-Late Flights



The estimated difference ranges from 0.37 to 0.89 fewer minutes per hour for very late departures. The p-value is extremely small, indicating a statistically significant difference between these groups.

Problem 4

Domain Problem: Does the average gain per hour differ for longer flights versus shorter flights?

Result Interpretation:

- Null Hypothesis (H_0):
  - Mean GPH of Long Flights *equals* Mean GPH of Short Flights
  - H_0: *μlong = μshort*

- Alternative Hypothesis (H_1):
  - Mean GPH of Long Flights *does not equal* Mean GPH of Short Flights
  - H_1: *μlong != μshort*

- Mean Gain Per Hour (GPH):
  - Long flights: 1.64 average gained minutes
  - Short flights: 5.59 average gained minutes

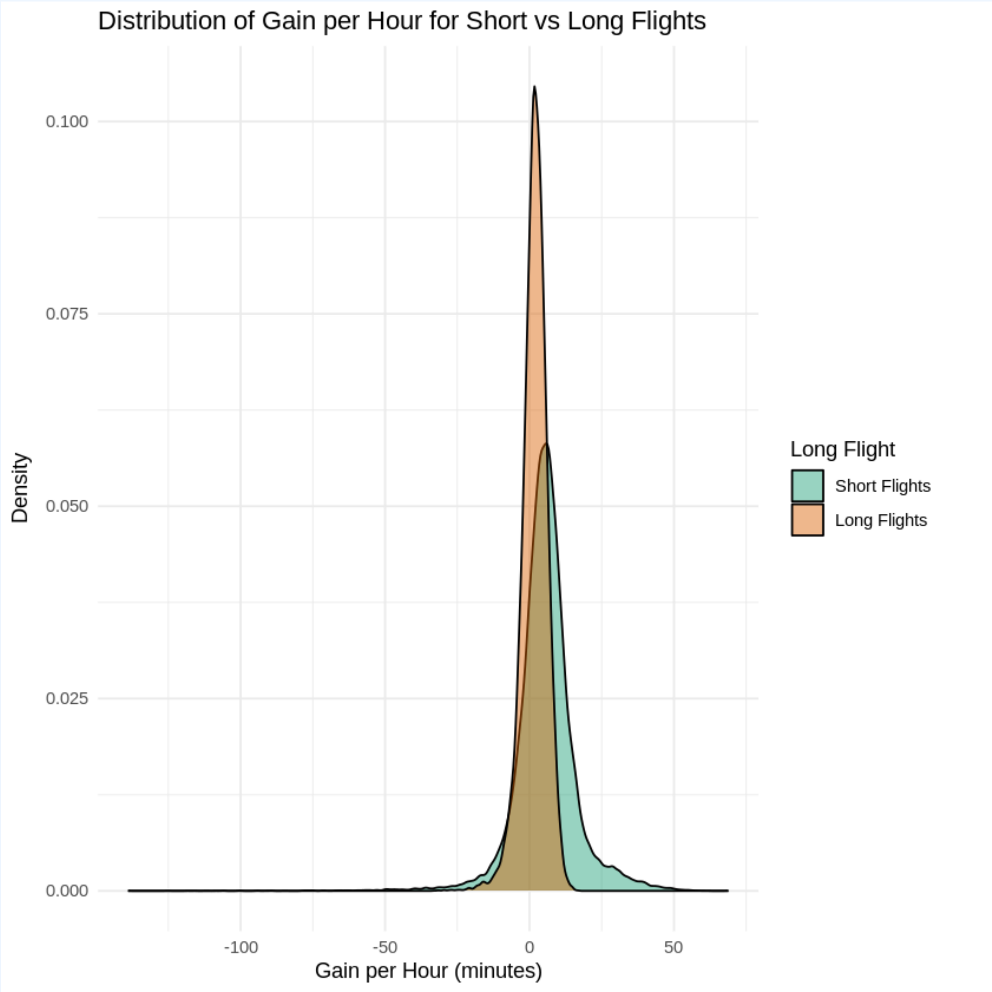
- Welch Two Sample T-Test Results:
  - t-statistic = 57.445
  - p-value < 2.2e-16
  - 95% confidence interval for the difference = [3.822, 4.0925]

For the hypothesis testing, we will follow standard practice:

- If p-value < 0.05, we reject H_0 and conclude that average gain per hour is significantly different between long and short flights.

- If p-value ≥ 0.05, we fail to reject H_0 and conclude that there is no strong evidence of a difference.

Distribution of Gain per Hour for Short vs Long Flights:



From the computed group means, we can observe that longer flights tend to have *less gain per hour*, with more flights departing from NYC; while shorter flights tend to have more gain *per hour*, with less flights departing from NYC. Overall, short flights gain much more time per hour than long flights — about 4 minutes more per hour, on average.

From the computed 95% confidence interval, we can more precisely determine the confidence levels in this observation. We are 95% confident that short flights gain between 3.8 and 4.1 more minutes per hour than long flights, and that this observation is statistically meaningful due to its positive values.

From our computed t-test, our p-value was less than 0.05. As a result, we reject the null hypothesis and conclude that there is extremely strong evidence to support that short flights and long flights differ in their average gain per hour.

KEY TAKEAWAYS

Finding 1: Late departures recover less time than on-time departures

Our analysis reveals that flights departing late gain 1.4-2.0 fewer minutes in the air compared to on-time flights. While this difference is statistically significant, the practical implications warrant careful consideration. For a typical 3-hour flight, this translates to only about 1% less recovery efficiency—a modest but consistent pattern across thousands of flights.

Why this matters for United Airlines: This finding challenges the assumption that pilots can simply "make up time" after ground delays. United should consider:

- **Schedule padding**: Building in appropriate buffer time on routes with frequent delays, particularly to/from congested airports
- **Crew training**: Understanding whether this limitation is due to air traffic control constraints, fuel optimization protocols, or other operational factors
- **Passenger communication**: Setting realistic arrival expectations when departures are delayed

Potential explanations: Late departures may have less flexibility to recover time due to:

- Congested air traffic patterns during peak delay periods
- Fuel conservation protocols that limit speed increases
- Cascading delays that affect optimal routing
- Crew duty time limitations that prevent extended flight times

Finding 2: The 30-minute threshold (very late flights) shows compounding effects

Flights delayed more than 30 minutes show even less recovery (1.3-2.2 fewer minutes), suggesting that severe delays may trigger operational constraints that further limit time recovery.

Operational insight: The 30-minute mark appears to be a critical threshold where United's ability to mitigate delays deteriorates. This could inform:

- **Proactive intervention protocols**: Implementing special procedures when delays approach 30 minutes
- **Resource allocation**: Prioritizing recovery efforts before this threshold is reached
- **Cancellation decisions**: Better cost-benefit analysis of when to cancel versus delay

Finding 3: Short flights show superior recovery efficiency

Our most striking finding is that short flights gain approximately 4 more minutes per hour than long flights (5.59 vs 1.64 minutes per hour). This 3-4x difference has significant operational implications.

Why short flights may recover better:

- More flexibility in cruise altitude adjustments
- Less complex routing with fewer waypoints
- Typically fly in less congested airspace

Strategic recommendations:

- **Route-specific strategies**: United should develop different recovery protocols for short-haul vs long-haul flights
- **Schedule design**: Short-haul routes may need less buffer time since they have greater recovery potential
- **Hub operations**: Since many short flights connect to long-haul flights, understanding this recovery differential is critical for minimizing missed connections

Limitations and Challenges

Data scope constraints: This analysis examines only 2013 data from a single carrier. The aviation industry has evolved significantly since then—changes in air traffic management technology, fuel efficiency protocols, and operational procedures may have altered recovery dynamics. Additionally, 2013 may not represent typical conditions if there were unusual weather patterns or industry disruptions that year.

Unmeasured confounding factors: Several variables that likely influence time recovery were not analyzed in our domain problems:

- **Weather conditions**: Headwinds, tailwinds, and storm systems dramatically affect flight times but weren't included in our model
- **Aircraft type**: Different aircraft have different cruise speeds and fuel efficiencies, affecting recovery capability
- **Pilot experience and behavior**: Individual pilot decisions about speed, altitude, and routing vary
- **Air traffic control directives**: We cannot distinguish between delays caused by airline decisions versus ATC requirements
- **Time of day**: Morning flights may face different air traffic patterns than evening flights

Challenges encountered:

During our analysis, we defined thresholds for "late" and "very late" flights using the 30-minute cutoff.

While commonly used in industry reporting, this threshold is somewhat arbitrary. Different thresholds, such as 15, 45, or 60 minutes, might reveal different patterns.

Additionally, deciding how to categorize short vs. long flights required careful consideration—using median split was intentional and practical but may mask more nuanced relationships.