

Thisisafunnygroupname's Project Report

Richard Zhou, Adam Rui, Jonathan Darius, Ojasvi Godha, Ryan Huang, Isaac Kang

Contents

Introduction	2
Project Description	2
Data Description/Preparation	2
Research Questions	3
Have flight delays improved over time overall? What about with individual airlines?	3
Do busy destinations tend to have more or less delays?	16
How does weather impact flight performance, and has it changed over time?	20
Does time of year affect flight delays?	29
Which airlines have the least delays? How has this changed over time?	33
Conclusions	38
Authors' Contributions	40
Data Availability	40
Links to Datasets	40

Introduction

Flight delays are a constant challenge in the air travel industry, impacting efficiency and passenger satisfaction. This project aims to investigate the underlying causes of flight delays in New York City and how these patterns have evolved over time. By analyzing both recent and historical flight data, we seek to identify the major contributors to delays and provide actionable insights for improving airline performance and the overall passenger experience.

Project Description

This analysis will utilize the `nycflights13` and `nycflights23` datasets, which contain records of flights departing from NYC airports. The project will involve exploratory data analysis (EDA), statistical testing, and comparative analysis, using tools such as `dplyr`, `ggplot2`, and many more to assess the significance of delay-related factors. Through this project, we intend to discover trends and patterns in flight delays, to provide a deeper insight into the aspects we can improve in air travel.

Through this data analysis, we aim to answer the 5 following questions:

1. Have flight delays improved over time overall?
 - What about with individual airlines?
2. Do busy destinations tend to have more or less delays?
3. Is the weather correlated with flight delays?
 - How has this changed over time?
4. Is the time of the year correlated between flight delays (holidays or rainy season)?
5. Which airlines have the least delays?
 - How has this changed over time?

Data Description/Preparation

This project utilizes data from the `nycflights13` and `nycflights23` packages, which provide detailed records of flights departing from New York City in 2013 and 2023, respectively. The two flight datasets were combined into a single data frame, with a new `year` variable added to distinguish between the two time periods. The resulting dataset, `flights_combined`, is also filtered to remove flights with missing departure or arrival delay values, and is further joined with the `airlines` and the `airports` datasets using appropriate keys. Our key attributes for this project include `dep_delay`, `arr_delay`, `carrier`, `dest`, and `year`, which are all key in exploring trends and changes in flight performance over the ten year period. We also used the `weather` dataset from both packages in order to find key relationships between delay patterns and weather.

Research Questions

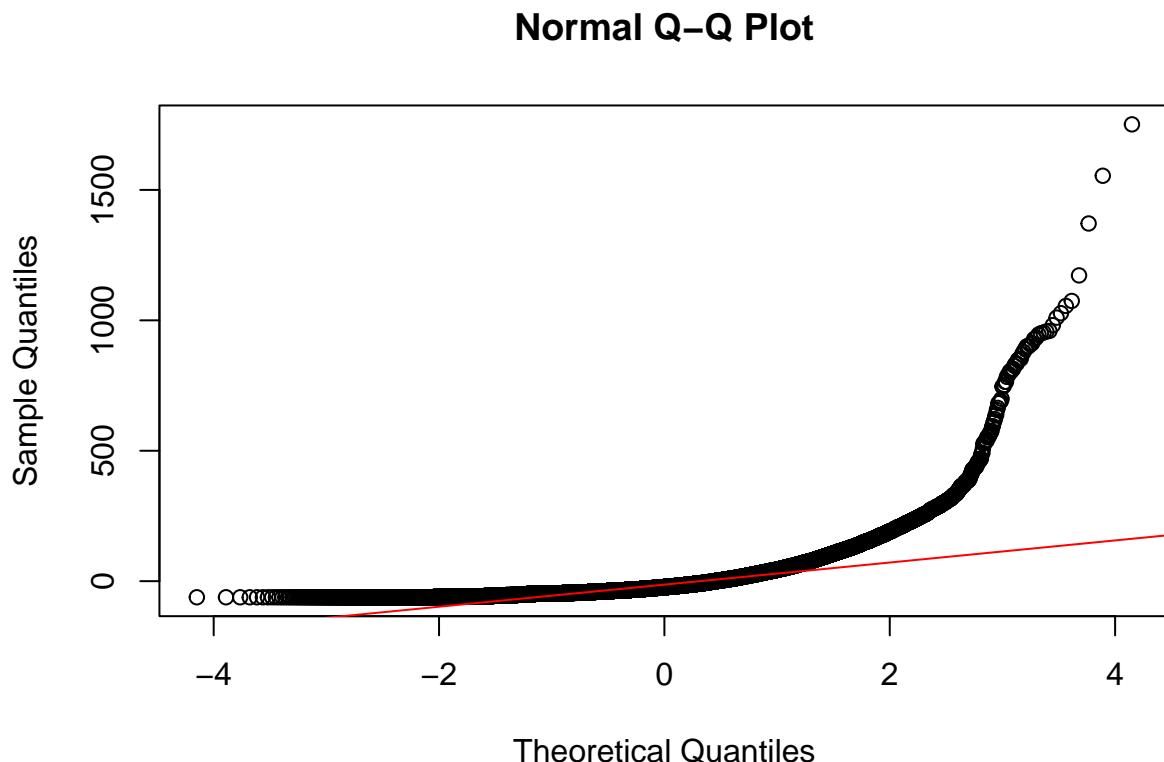
Have flight delays improved over time overall? What about with individual airlines?

Part 1: Have flight delays improved or gotten worse between 2013 and 2023?

First we test for normality.

```
dep_delay_resids <- sample(residuals(dep_delay_model), size = 30000)

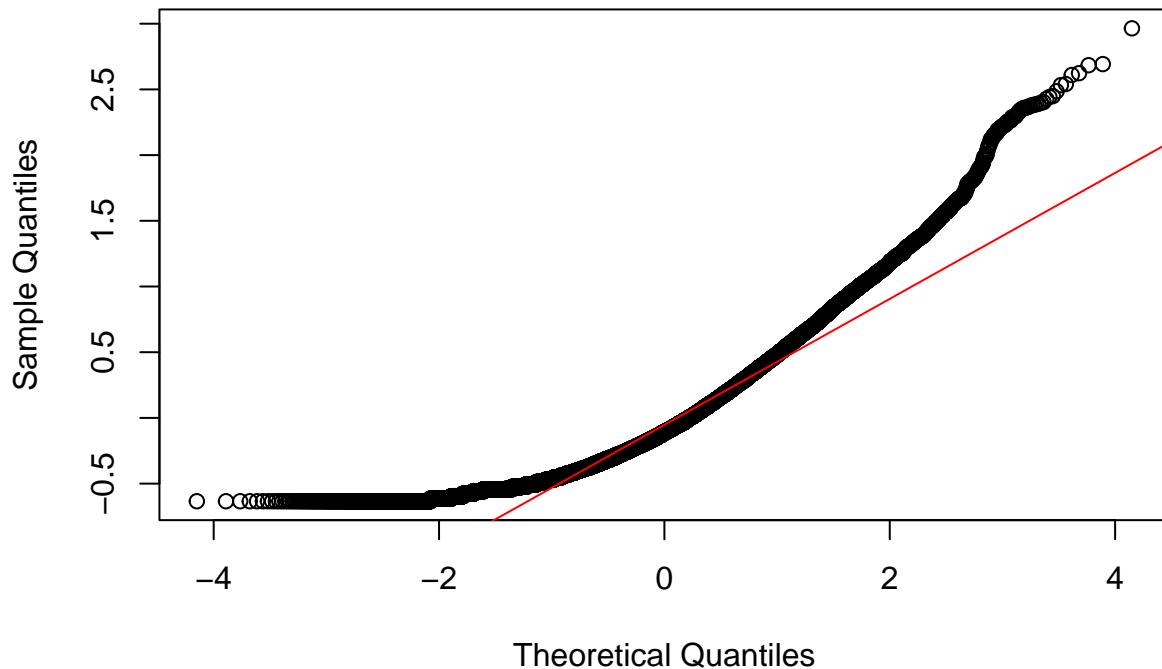
# Now plot the Q-Q plot with the sample for easier loading
qqnorm(dep_delay_resids)
qqline(dep_delay_resids, col = "red")
```



We don't seem to meet this assumption so let's do a shifted log transformation to help. This type of log transformation ensures all values are positive before the logging takes place and makes sure to include all values possible. After, we are going to check the normality assumption again.

```
# Now plot the Q-Q plot with the sample for easier loading
qqnorm(dep_delay_log_resids)
qqline(dep_delay_log_resids, col = "red")
```

Normal Q-Q Plot



And since the points are fairly close to the red line, we can say that the data is not from a perfect normal distribution, so despite this we will continue as it is fairly close.

```
leveneTest(log_dep_delay ~ factor(year), data = flights_clean_log)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      1 608.81 < 2.2e-16 ***
##        210624
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can also check for homoscedasticity using the Levene's Test which is not met. The small p-value here shows signs of heteroscedasticity which means we cannot use regular 'summary()' to get results of the model. Let's look into the model itself using robust standard errors which assumes unequal error variances.

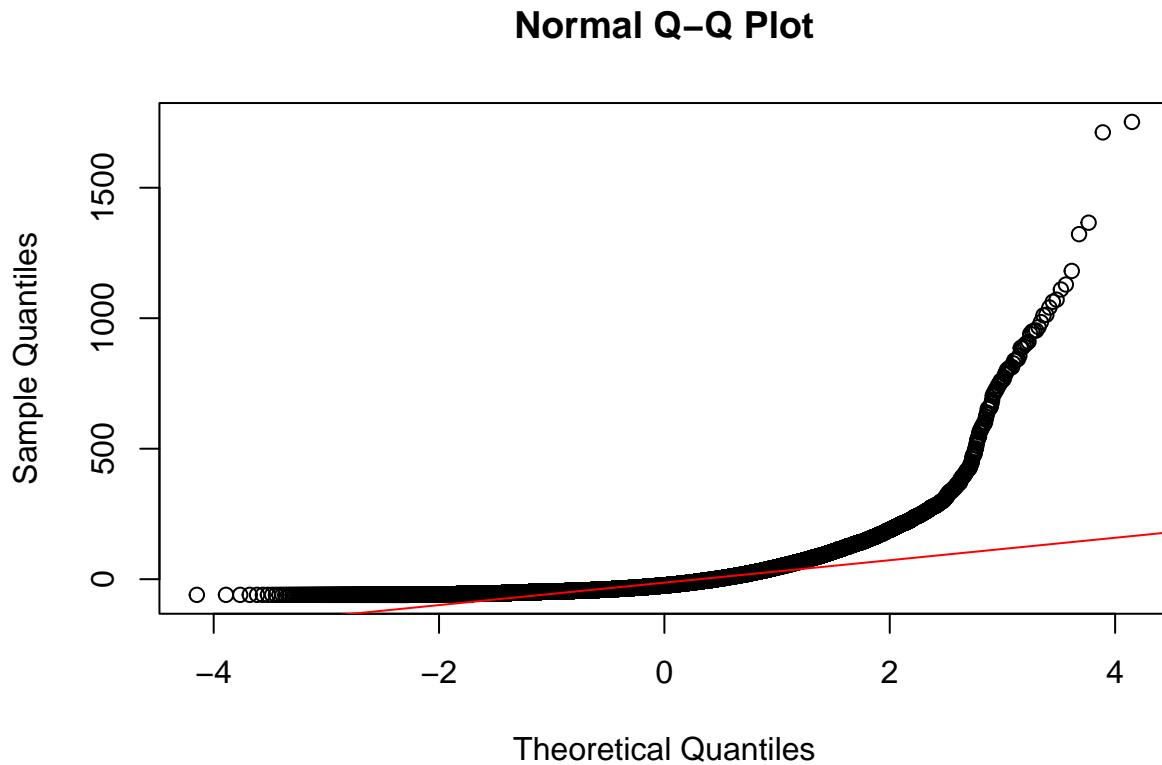
```
coeftest(dep_delay_model_log, vcov = vcovHC)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.4764619  0.0014273 3136.424 < 2.2e-16 ***
## factor(year)2023 0.0894405  0.0020859   42.879 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This model helps us see the overall difference in departure delays between 2013 and 2023. The model shows a statistically significant increase in departure delays from 2013 to 2023, indicated by a positive coefficient with a very small p-value. In this context, we can conclude that flight departure delays have gotten worse over time.

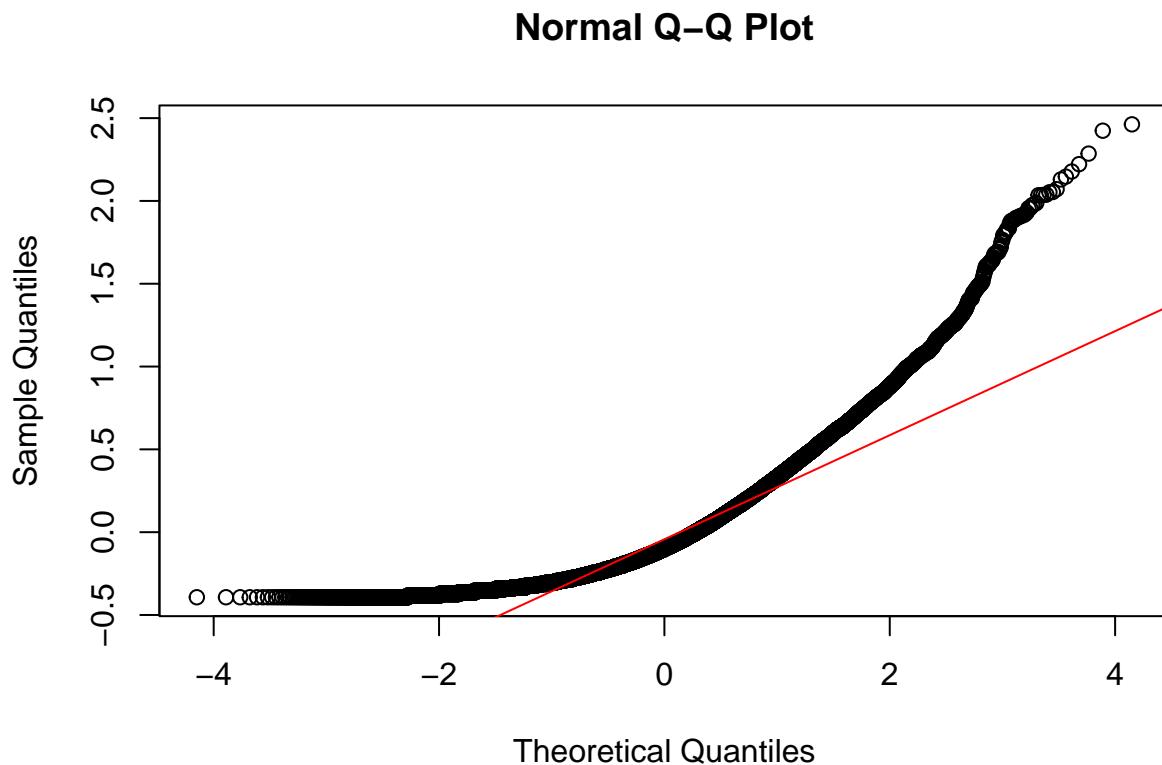
Let's do the whole process again, with arrival delays.

```
# Now plot the Q-Q plot with the sample for easier loading
qqnorm(arr_delay_resids)
qqline(arr_delay_resids, col = "red")
```



We seem to have the same issue as the departure delays, so let's do another shifted log transformation and test again.

```
# Now plot the Q-Q plot with the sample for easier loading
qqnorm(arr_delay_log_resids)
qqline(arr_delay_log_resids, col = "red")
```



Again, the points are fairly close to the red line, so we can say that the data is not from a perfect normal distribution, it is close enough to continue.

```
leveneTest(log_arr_delay ~ factor(year), data = flights_clean_log)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      1 935.59 < 2.2e-16 ***
##          210624
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can, again, check for homoscedasticity using the Levene's Test which is, again, not met. The small p-value here shows signs of heteroscedasticity which means we cannot use regular 'summary()' to get results of the model. Let's look into the model itself using robust standard errors which assumes unequal error variances.

```
coeftest(arr_delay_model_log, vcov = vcovHC)
```

```
##
## t test of coefficients:
##
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.93679104 0.00098648 5004.466 < 2.2e-16 ***
## factor(year)2023 0.04099626 0.00148077   27.686 < 2.2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When we compare the arrival delays, we have a statistically significant increase as well, with a positive coefficient and small p-value, indicating that arrival delays have also gotten worse between 2013 and 2023, however slightly.

Let's take a quick look at the performance of both these arrival and delay models.

```
print(paste('Arrival Adj R^2: ', summary(arr_delay_model_log)$adj.r.squared))
## [1] "Arrival Adj R^2:  0.0035573193253432"

print(paste("Departure Adj R^2: ", summary(dep_delay_model_log)$adj.r.squared))
## [1] "Departure Adj R^2:  0.00854136653743243"
```

The Adjusted R² value helps us measure the quality of the model. These are really small values which means these models are not great. But in this case the small values are fine as the dataset is large, and the focus of this question was to identify average differences in delays over time, so the models still provided meaningful insights for this question.

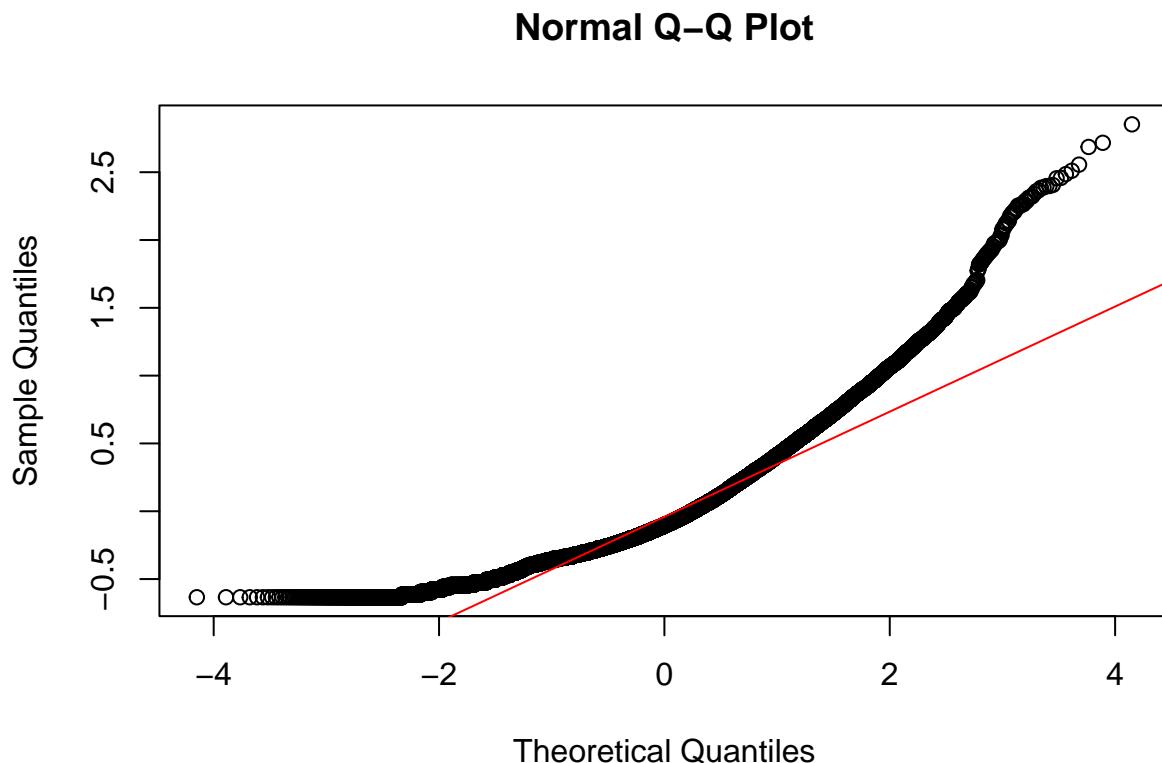
For some additional comparison, we can try to see how much more or less departure delays have changed versus arrival delays from 2013 to 2023.

```
both_delay_flights <- flights_clean_log %>%
  select(year, log_dep_delay, log_arr_delay) %>%
  pivot_longer(cols = c(log_dep_delay, log_arr_delay),
               names_to = "delay_type",
               values_to = "delay_value")

both_delay_model <- lm(data=both_delay_flights, delay_value~factor(year)*delay_type)
```

Let's first check for the normality assumption.

```
# Now plot the Q-Q plot with the sample for easier loading
qqnorm(both_delay_resids)
qqline(both_delay_resids, col = "red")
```



Similar to plots before, we can say we have enough to continue despite not perfectly meeting the normality assumption. Now let's check for homoscedasticity with the Levene's Test.

```
leveneTest(delay_value ~ factor(year)*delay_type, data = both_delay_flights)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      3 6481 < 2.2e-16 ***
##             421248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can check for homoscedasticity using the Levene's Test which is, again, not met. The small p-value here shows signs of heteroscedasticity which means we cannot use regular `'summary()'` to get results of the model. Let's look into the model itself using robust standard errors which assumes unequal error variances.

```
coeftest(both_delay_model, vcov = vcovHC)
```

```
##
## t test of coefficients:
##
##                               Estimate Std. Error t value
## (Intercept)                4.93679104  0.00098648 5004.466
## factor(year)2023            0.04099626  0.00148077   27.686
## delay_type log_dep_delay -0.46032914  0.00173499 -265.321
```

```

## factor(year)2023:delay_type log_dep_delay 0.04844426 0.00255802 18.938
##                                     Pr(>|t|)
## (Intercept) < 2.2e-16 ***
## factor(year)2023 < 2.2e-16 ***
## delay_type log_dep_delay < 2.2e-16 ***
## factor(year)2023:delay_type log_dep_delay < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The key term in the model is ‘factor(year)2023:delay_type log_dep_delay’, which represents the additional change in departure delays from 2013 to 2023 relative to arrival delays. From the two previous models, we know arrival delays increased over the years by about 10% and departure delays increased by about 28%. In this combined model, we can see that as the coefficient is positive and, based on the p-value, is statistically significant. With this, we can conclude that the departure delays not only increased over time, but did so to a significantly greater extent than arrival delays did.

Part 2: Have individual airlines gotten better or worse with delays over time?

```

delay_model_airline <- lm(data=flights_clean_log, log_dep_delay~factor(year) * name.x)
summary(delay_model_airline)

```

```

##
## Call:
## lm(formula = log_dep_delay ~ factor(year) * name.x, data = flights_clean_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7388 -0.3684 -0.1064  0.2664  2.9494
##
## Coefficients: (5 not defined because of singularities)
##                                         Estimate Std. Error t value
## (Intercept)                         4.412264  0.012409 355.577
## factor(year)2023                      0.139575  0.004658 29.962
## name.xAlaska Airlines Inc.            0.070165  0.043542  1.611
## name.xAmerican Airlines Inc.          0.060822  0.013616  4.467
## name.xDelta Air Lines Inc.           0.040264  0.013206  3.049
## name.xEndeavor Air Inc.              0.182796  0.014031 13.028
## name.xEnvoy Air                     0.050477  0.013550  3.725
## name.xExpressJet Airlines Inc.        0.153855  0.012859 11.965
## name.xFrontier Airlines Inc.          0.057058  0.030807  1.852
## name.xHawaiian Airlines Inc.          0.001792  0.075629  0.024
## name.xJetBlue Airways                0.053297  0.012916  4.126
## name.xMesa Airlines Inc.              0.159252  0.035445  4.493
## name.xSkyWest Airlines Inc.            0.236242  0.169346  1.395
## name.xSouthwest Airlines Co.          0.038806  0.014257  2.722
## name.xUnited Air Lines Inc.           -0.002563  0.012907 -0.199
## name.xUS Airways Inc.                 -0.033307  0.014476 -2.301
## name.xVirgin America                  0.037862  0.017907  2.114
## factor(year)2023:name.xAlaska Airlines Inc. -0.066792  0.043460 -1.537
## factor(year)2023:name.xAmerican Airlines Inc. -0.031584  0.008813 -3.584
## factor(year)2023:name.xDelta Air Lines Inc.    -0.022266  0.007576 -2.939
## factor(year)2023:name.xEndeavor Air Inc.     -0.196119  0.009324 -21.034

```

```

## factor(year)2023:name.xEnvoy Air           -0.179610  0.046106 -3.896
## factor(year)2023:name.xExpressJet Airlines Inc.      NA        NA       NA
## factor(year)2023:name.xFrontier Airlines Inc.    0.061739  0.034918  1.768
## factor(year)2023:name.xHawaiian Airlines Inc.   -0.160427  0.082629 -1.942
## factor(year)2023:name.xJetBlue Airways        0.026472  0.006633  3.991
## factor(year)2023:name.xMesa Airlines Inc.       NA        NA       NA
## factor(year)2023:name.xSkyWest Airlines Inc.   -0.170036  0.169349 -1.004
## factor(year)2023:name.xSouthwest Airlines Co.  -0.174678  0.011056 -15.799
## factor(year)2023:name.xUnited Air Lines Inc.     NA        NA       NA
## factor(year)2023:name.xUS Airways Inc.          NA        NA       NA
## factor(year)2023:name.xVirgin America         NA        NA       NA
##
##                                         Pr(>|t|)
## (Intercept)                         < 2e-16 ***
## factor(year)2023                      < 2e-16 ***
## name.xAlaska Airlines Inc.            0.107088
## name.xAmerican Airlines Inc.         7.94e-06 ***
## name.xDelta Air Lines Inc.           0.002297 **
## name.xEndeavor Air Inc.              < 2e-16 ***
## name.xEnvoy Air                     0.000195 ***
## name.xExpressJet Airlines Inc.       < 2e-16 ***
## name.xFrontier Airlines Inc.         0.064014 .
## name.xHawaiian Airlines Inc.         0.981099
## name.xJetBlue Airways               3.69e-05 ***
## name.xMesa Airlines Inc.             7.03e-06 ***
## name.xSkyWest Airlines Inc.          0.163011
## name.xSouthwest Airlines Co.        0.006494 **
## name.xUnited Air Lines Inc.         0.842615
## name.xUS Airways Inc.               0.021402 *
## name.xVirgin America                0.034485 *
## factor(year)2023:name.xAlaska Airlines Inc. 0.124328
## factor(year)2023:name.xAmerican Airlines Inc. 0.000339 ***
## factor(year)2023:name.xDelta Air Lines Inc.  0.003292 **
## factor(year)2023:name.xEndeavor Air Inc.    < 2e-16 ***
## factor(year)2023:name.xEnvoy Air        9.80e-05 ***
## factor(year)2023:name.xExpressJet Airlines Inc.  NA
## factor(year)2023:name.xFrontier Airlines Inc.  0.077041 .
## factor(year)2023:name.xHawaiian Airlines Inc.  0.052195 .
## factor(year)2023:name.xJetBlue Airways      6.59e-05 ***
## factor(year)2023:name.xMesa Airlines Inc.     NA
## factor(year)2023:name.xSkyWest Airlines Inc.  0.315354
## factor(year)2023:name.xSouthwest Airlines Co. < 2e-16 ***
## factor(year)2023:name.xUnited Air Lines Inc.  NA
## factor(year)2023:name.xUS Airways Inc.        NA
## factor(year)2023:name.xVirgin America        NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## Residual standard error: 0.4777 on 192181 degrees of freedom
##   (18418 observations deleted due to missingness)
## Multiple R-squared:  0.02332,   Adjusted R-squared:  0.02319
## F-statistic: 176.5 on 26 and 192181 DF,  p-value: < 2.2e-16

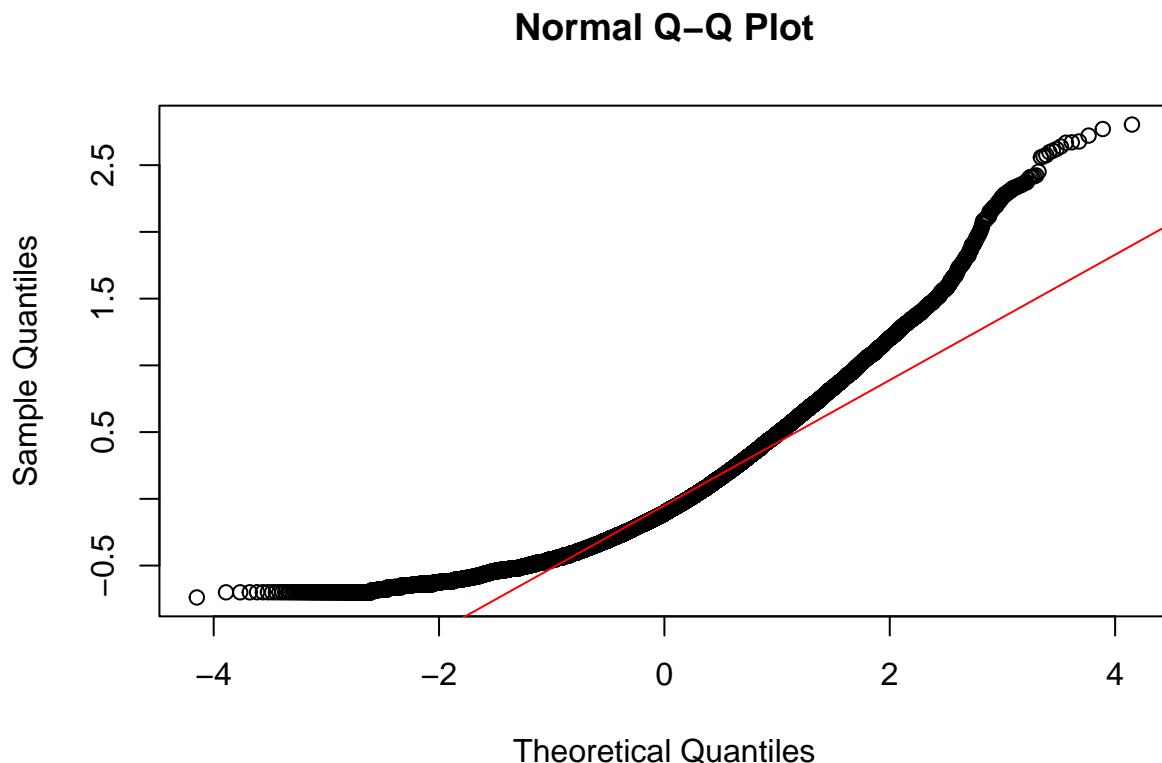
```

Based on the F-statistic, this model is significant. However, there are a few airlines that seem to be discontinued in 2023, so lets remove them and create a new model.

```
dep_delay_model_airline_filtered <- lm(data=flights_filtered, log_dep_delay~factor(year) * name.x)
```

Now with our new model, let's check for the normality assumption.

```
# Now plot the Q-Q plot with the sample for easier loading
qqnorm(dep_delay_airline_resids)
qqline(dep_delay_airline_resids, col = "red")
```



Despite the skewness suggesting we don't perfectly meet the normality assumption, we will continue. Now let's check for homoscedasticity with the Levene's Test.

```
leveneTest(log_dep_delay ~ factor(year) * name.x, data = flights_filtered)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      21 78.152 < 2.2e-16 ***
##          164960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the small p-value, we can see signs of heteroscedasticity which means we cannot use regular 'summary()' to get results of the model. Let's look into the model itself using robust standard errors which assumes unequal error variances.

```

dep_delay_robust_se <- vcovHC(dep_delay_model_airline_filtered, type = "HC1")
dep_delay_tidy_robust_model <- tidy(coeftest(dep_delay_model_airline_filtered, vcov. = dep_delay_robust_se))

# only getting interaction terms, what we need
dep_interaction_terms <- dep_delay_tidy_robust_model[grep(":", dep_delay_tidy_robust_model$term),]
print(dep_interaction_terms)

## # A tibble: 10 x 5
##   term                      estimate std.error statistic p.value
##   <chr>                    <dbl>    <dbl>     <dbl>    <dbl>
## 1 factor(year)2023:name.xAmerican Airline~  0.0352    0.0432    0.816   0.415
## 2 factor(year)2023:name.xDelta Air Lines ~  0.0445    0.0428    1.04    0.299
## 3 factor(year)2023:name.xEndeavor Air Inc. -0.129    0.0431   -3.00    0.00271
## 4 factor(year)2023:name.xEnvoy Air        -0.113    0.0525   -2.15    0.0317
## 5 factor(year)2023:name.xFrontier Airline~  0.129    0.0563    2.28    0.0225
## 6 factor(year)2023:name.xHawaiian Airline~ -0.0936   0.117    -0.802   0.422
## 7 factor(year)2023:name.xJetBlue Airways   0.0933   0.0426    2.19    0.0287
## 8 factor(year)2023:name.xSkyWest Airlines~ -0.103    0.168    -0.614   0.539
## 9 factor(year)2023:name.xSouthwest Airlin~ -0.108    0.0434   -2.49    0.0129
## 10 factor(year)2023:name.xUnited Air Lines~  0.0668   0.0426    1.57    0.117

```

Based on the p-values, only 3 of these 10 airlines are statistically significant. Yet despite that we can still gain an idea of a general trend by looking at the coefficients. 6 out of 10 of these coefficients are positive, which means the majority of the airlines have gotten worse in 2023 compared to 2013.

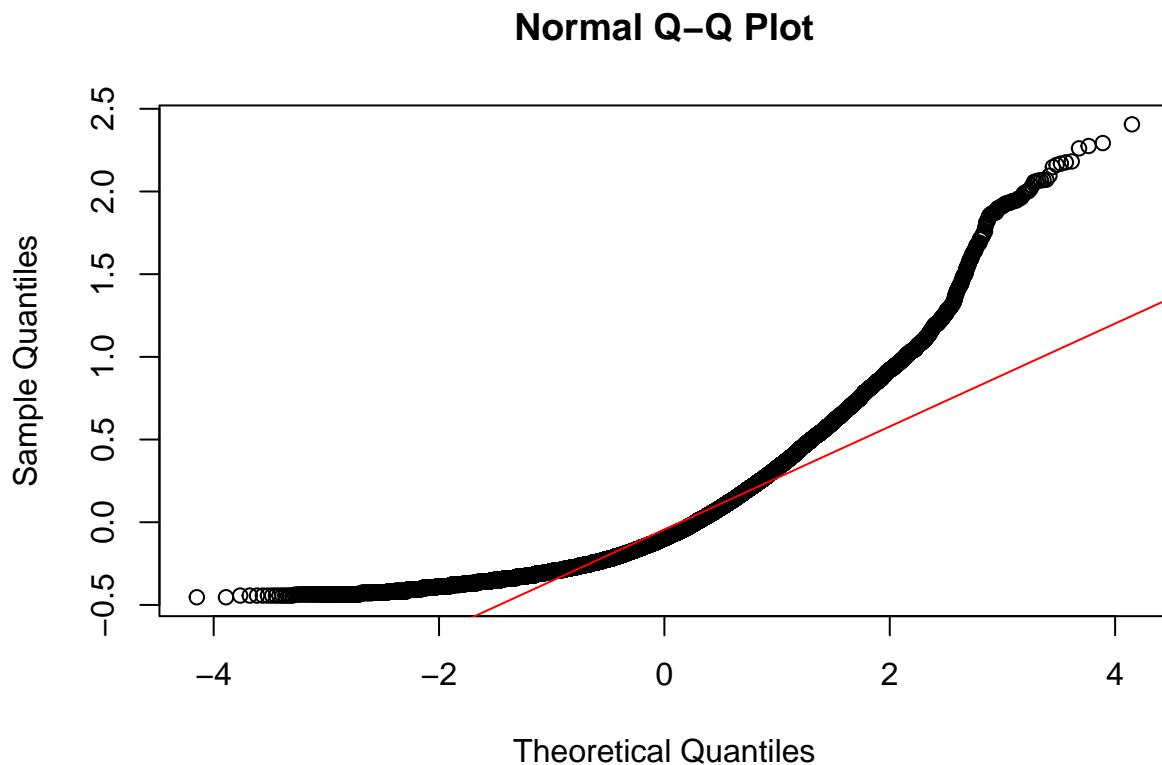
Let's do this again with arrival delays.

Let's check for the normality assumption.

```

# Now plot the Q-Q plot with the sample for easier loading
qqnorm(arr_delay_airline_resids)
qqline(arr_delay_airline_resids, col = "red")

```



Based on this plot, despite the slight skewness, we can say we meet the normality assumption in a manner similar to previous plots. Now let's check for homoscedasticity with the Levene's Test.

```
leveneTest(log_arr_delay ~ factor(year) * name.x, data = flights_filtered)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      21 89.975 < 2.2e-16 ***
##             164960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the small p-value, we can see signs of heteroscedasticity which means we cannot use regular 'summary()' to get results of the model. Let's look into the model itself using robust standard errors which assumes unequal error variances.

```
arr_delay_robust_se <- vcovHC(arr_delay_model_airline_filtered, type = "HC1")
arr_delay_tidy_robust_model <- tidy(coeftest(arr_delay_model_airline_filtered, vcov. = arr_delay_robust_se))

# only getting interaction terms, what we need
arr_interaction_terms <- arr_delay_tidy_robust_model[grep(":", arr_delay_tidy_robust_model$term),]
print(arr_interaction_terms)

## # A tibble: 10 x 5
##   term                estimate std.error statistic p.value
##   <chr>              <dbl>     <dbl>     <dbl>     <dbl>
```

```

##      <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 factor(year)2023:name.xAmerican Airline~  0.00770   0.0271   0.284  7.76e-1
## 2 factor(year)2023:name.xDelta Air Lines ~ -0.0189   0.0268  -0.706  4.80e-1
## 3 factor(year)2023:name.xEndeavor Air Inc. -0.102    0.0271  -3.78   1.58e-4
## 4 factor(year)2023:name.xEnvoy Air        -0.181    0.0332  -5.46   4.73e-8
## 5 factor(year)2023:name.xFrontier Airline~ -0.0161   0.0380  -0.425  6.71e-1
## 6 factor(year)2023:name.xHawaiian Airline~ -0.0663   0.0771  -0.860  3.90e-1
## 7 factor(year)2023:name.xJetBlue Airways   0.0172   0.0267   0.645  5.19e-1
## 8 factor(year)2023:name.xSkyWest Airlines~ -0.136    0.110   -1.24   2.17e-1
## 9 factor(year)2023:name.xSouthwest Airlin~ -0.108    0.0272  -3.95   7.69e-5
## 10 factor(year)2023:name.xUnited Air Lines~  0.00695   0.0267   0.261  7.94e-1

```

Based on the p-values here, again only 3 of these 10 airlines are statistically significant. We can still gain an idea of a general trend by looking at the coefficients. 6 out of 10 of these coefficients are negative, opposite of the departure trends. This means the majority of the airlines have actually gotten better in 2023 compared to 2013.

There are a few important limitations to note in this analysis. First, the assumption of independence may be violated, as flights from the same airline or airport are likely to be correlated. However, given the large sample size and the goal of identifying average differences in delays over time, we proceeded with the models, which still offered meaningful insights. Another limitation involves the normality assumption—although we applied shifted log transformations, the residuals still showed some skewness. Given the size of the dataset, we considered the approximation acceptable. With these limitations in mind, the findings remain useful, but should be interpreted with some caution.

Graphs

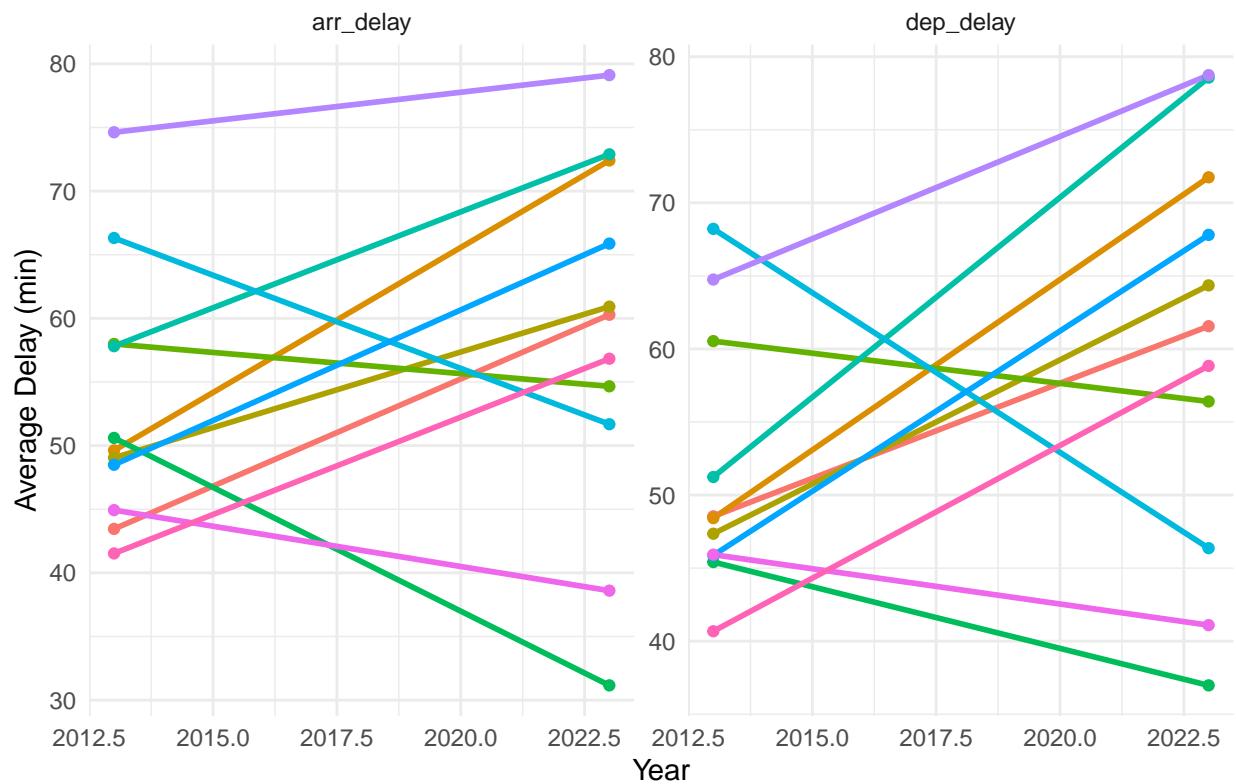
Here, we can easily see how the average arrival and departure delays have changed between 2013 and 2023. It seems that both arrival and departure delays have gotten worse.

```

## `summarise()` has grouped output by 'name.x', 'year'. You can override using
## the '.groups' argument.

```

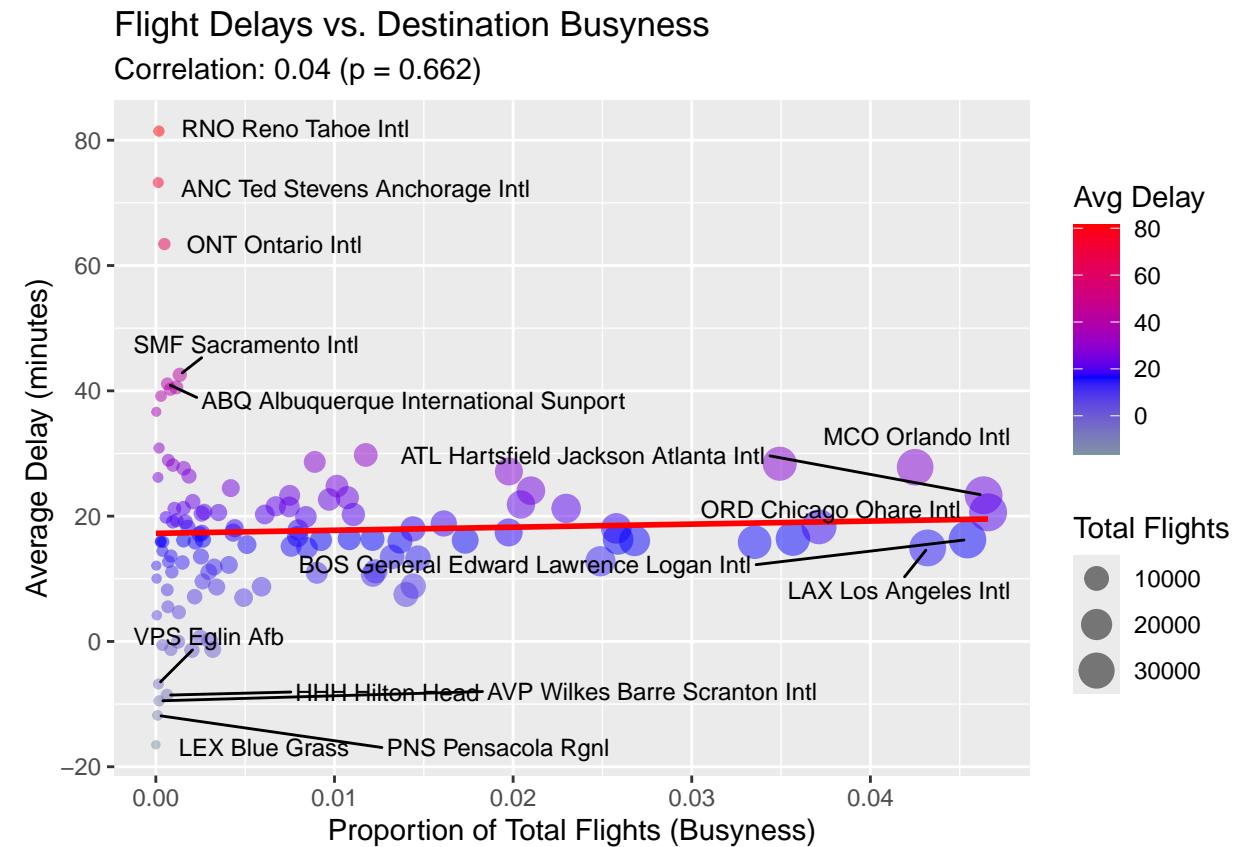
Trends in Delay by Airline (2013 vs 2023)



Do busy destinations tend to have more or less delays?

Data Exploration and Visualization

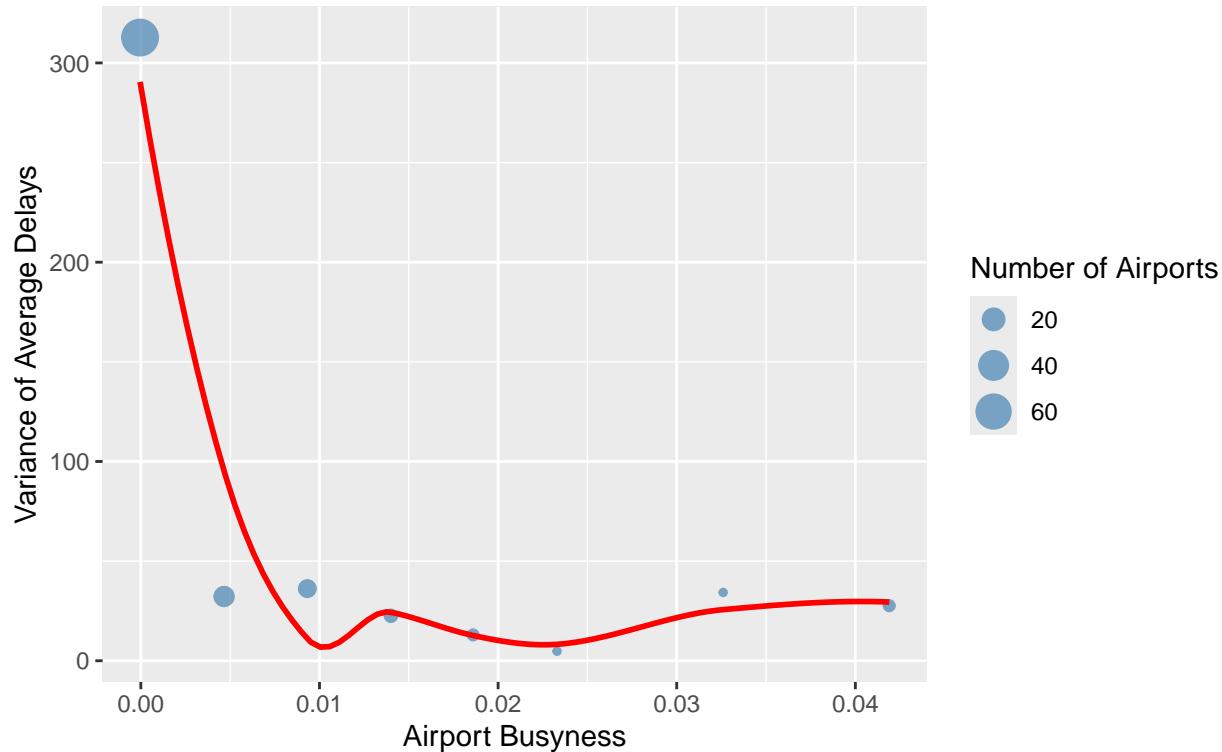
```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'
```

Heteroscedasticity Check: Variance of Delays vs. Busyness

Each point represents a group of airports with similar busyness levels



The scatter plot shows the average delay against the busyness of each airport, where the colour is also indicative of delay length, and the size of the point is also indicative of busyness. Using the simple linear fit demonstrates no statistical significance of correlation between the two variables, however assumption checking may reveal otherwise (it doesn't, but always check). It's also apparent that as busyness increases the average delay tends to stabilize (as demonstrated in the Heteroscedasticity graph) to around 20 minutes, potentially being correlated with the available infrastructure at each airport to minimize random issues that result in higher delays.

Data Analysis/Modeling/Predictions

```
model <- lm(avg_delay ~ busyness, data = destination_stats)
bptest(model) # p > 0.05 = homoscedastic
```

```
##
## studentized Breusch-Pagan test
##
## data: model
## BP = 5.4403, df = 1, p-value = 0.01968
```

The Breusch-Pagan test reveals that the data is not homoscedastic, with a p-value below 0.05 (0.0196), indicating that the variance isn't constant.

```

shapiro.test(residuals(model))

##
## Shapiro-Wilk normality test
##
## data: residuals(model)
## W = 0.86554, p-value = 6.72e-09

```

The Shapiro-wilk test reveals that the data is not normal, with a p value < 0.05 (6.72e-09).

```

# accounting for heteroscedasticity (robust standard error)
coeftest(model, vcov = vcovHC(model, type = "HC1"))

```

```

##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.2503   1.9068  9.0469 4.304e-15 ***
## busyness    49.0837  82.8622  0.5924   0.5548
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Using robust standard errors to account for heteroscedasticity, reveals there is still no statistical significance between the correlation of delay and busyness (p-value of 0.5548, > 0.05).

```

# accounting for normality (np regression)
model_gam <- gam(avg_delay ~ s(busyness), data = destination_stats)
summary(model_gam)

```

```

##
## Family: gaussian
## Link function: identity
##
## Formula:
## avg_delay ~ s(busyness)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.670     1.282    13.79  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df      F p-value
## s(busyness)  1       1 0.192  0.662
##
## R-sq.(adj) = -0.00701  Deviance explained = 0.167%
## GCV = 195.52  Scale est. = 192.17  n = 117

```

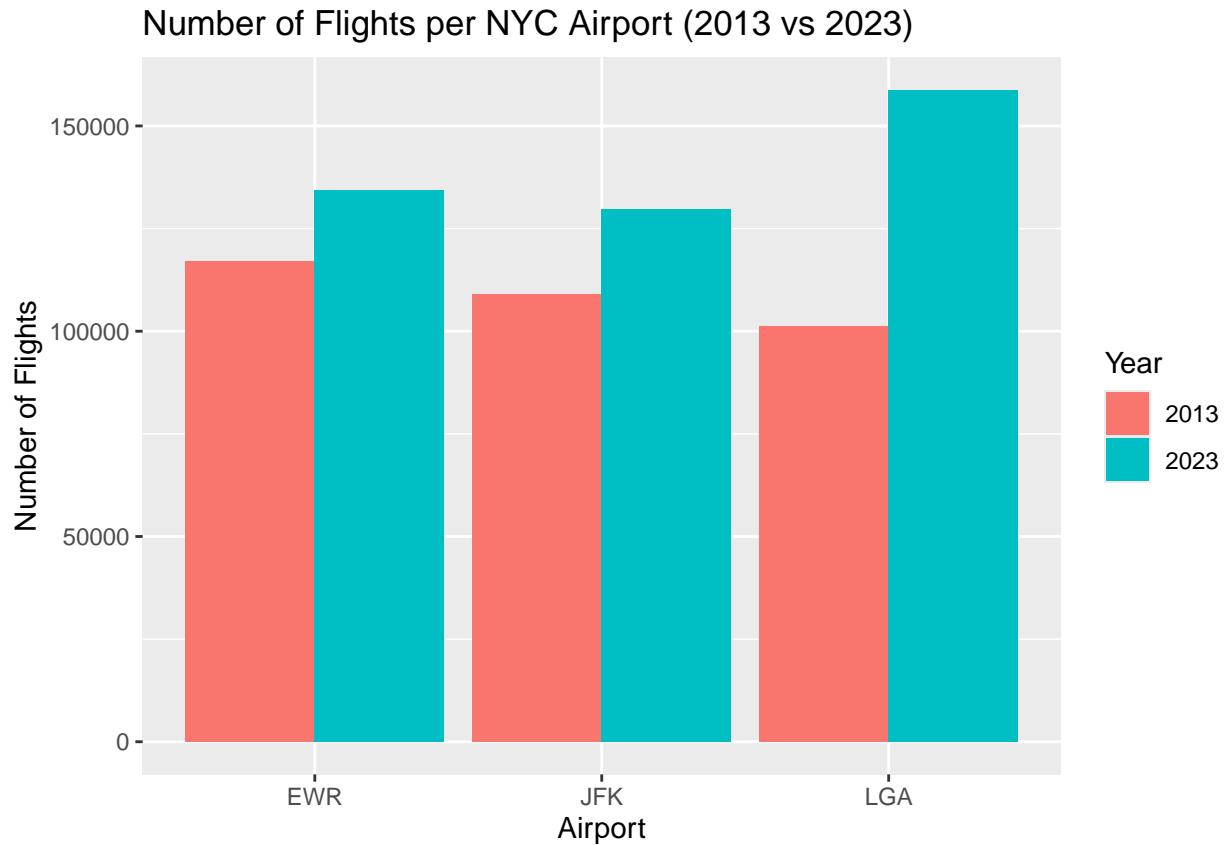
The GAM produced an smooth term p = 0.662, which indicates that the smooth term is not significant, which is backed up by the R^2 value of near zero (-0.00701), meaning almost none of the variance can be explained by busyness.

Results and Insights

Testing revealed that the data was neither normal, nor homoscedastic. Analysis revealed that busyness has no statistically significant correlation with flight delays ($p = 0.55$, robust SEs), with the model explaining virtually no variance (adj. $R^2 = -0.007$). A possible explanation for these issues is that delays may have a volume dependent variability, where busier airports can have both extremely on time flights and extremely delayed flights, increasing variance, or the large volume of small airports, where delays may be sporadic and unpredictable. It's also likely that most flights are on time, or have a very short delay, resulting in an extreme right skew, violating non normality. Several limitations of only analyzing busyness and delay, arise, such as the under fitting of the GAM, with an R^2 near zero, indicating that the other predictors are missing (although they weren't required for this analysis specifically). Furthermore, although we tried to address the assumption violations with robust standard errors, the underlying skew of the data will probably require other methods to account for. Another important factor that was not accounted for was congestion of flights, different from busyness, as due to airport / airspace design, a small volume of flights may overwhelm an airport, resulting in an airport instead of busyness based delay. Our null result is still significant, however, as it is conclusive that more flights is not equal to longer delays.

How does weather impact flight performance, and has it changed over time?

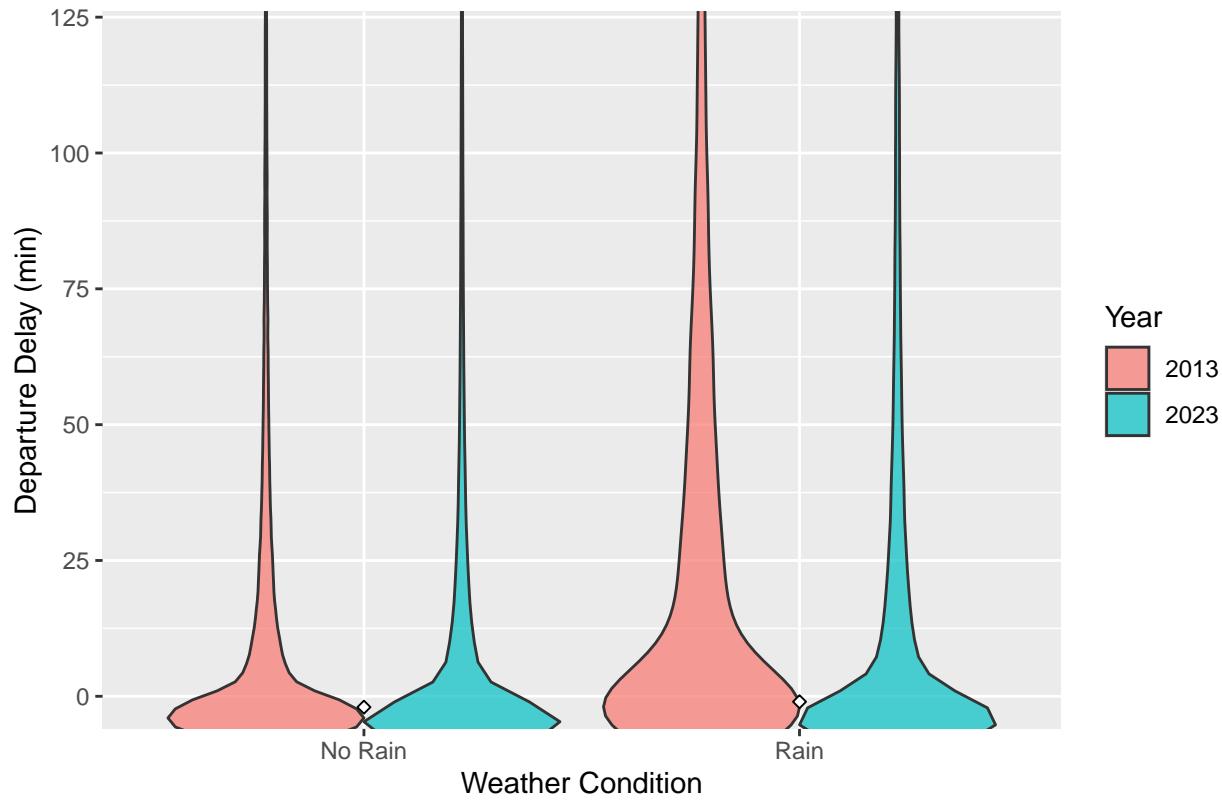
Data Exploration and Visualization



This bar chart compares the number of flights at NYC airports (EWR, JFK, and LGA) in 2013 and 2023. All three airports experienced an increase in flights, with LaGuardia (LGA) showing the most significant growth. This may reflect recent infrastructure improvements at LGA and a broader rise in domestic air travel. The visualization clearly highlights these trends by contrasting the flight volumes side by side. Notably, LGA's dramatic rise may suggest a shift in airline operations or passenger preference toward more centrally located airports.

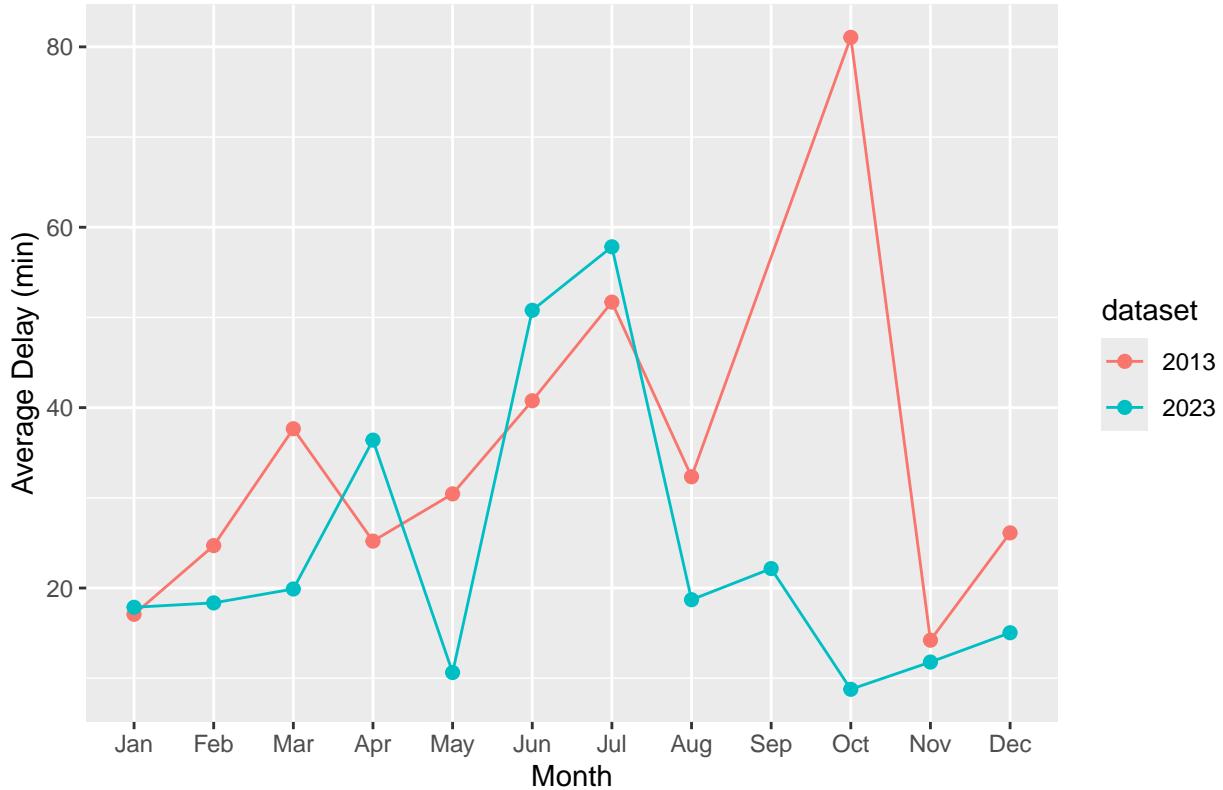
We want to combine data sets from both 2013 and 2023. If we combine these data sets, we can have a combined data frame that contains data from both years, which would allow us to visualize and compare conditions across both years.

Departure Delay with Rain vs No Rain



This violin plot compares departure delays under rainy and non-rainy conditions for the years 2013 and 2023. Overall, delays were higher during rain in both years, especially in 2013, suggesting weather remained a key factor in flight delays. Notably, 2023 shows a reduction in departure delays under both conditions, which may indicate improved scheduling, infrastructure, or weather forecasting technology. The visualization effectively shows not just the average delay but the distribution, revealing that extreme delays became less frequent by 2023.

Monthly Delay During Rain (2013 vs 2023)



This line plot compares average monthly departure delays during rain in 2013 and 2023. Overall, 2023 shows consistently lower delays, especially during peak months like October, suggesting improvements in weather-related operations. The sharp spike in delays during October 2013 highlights a possible extreme weather event or operational issue that year. The visualization helps reveal both seasonal patterns and long-term improvements in delay management.

```
## `geom_smooth()` using formula = 'y ~ x'

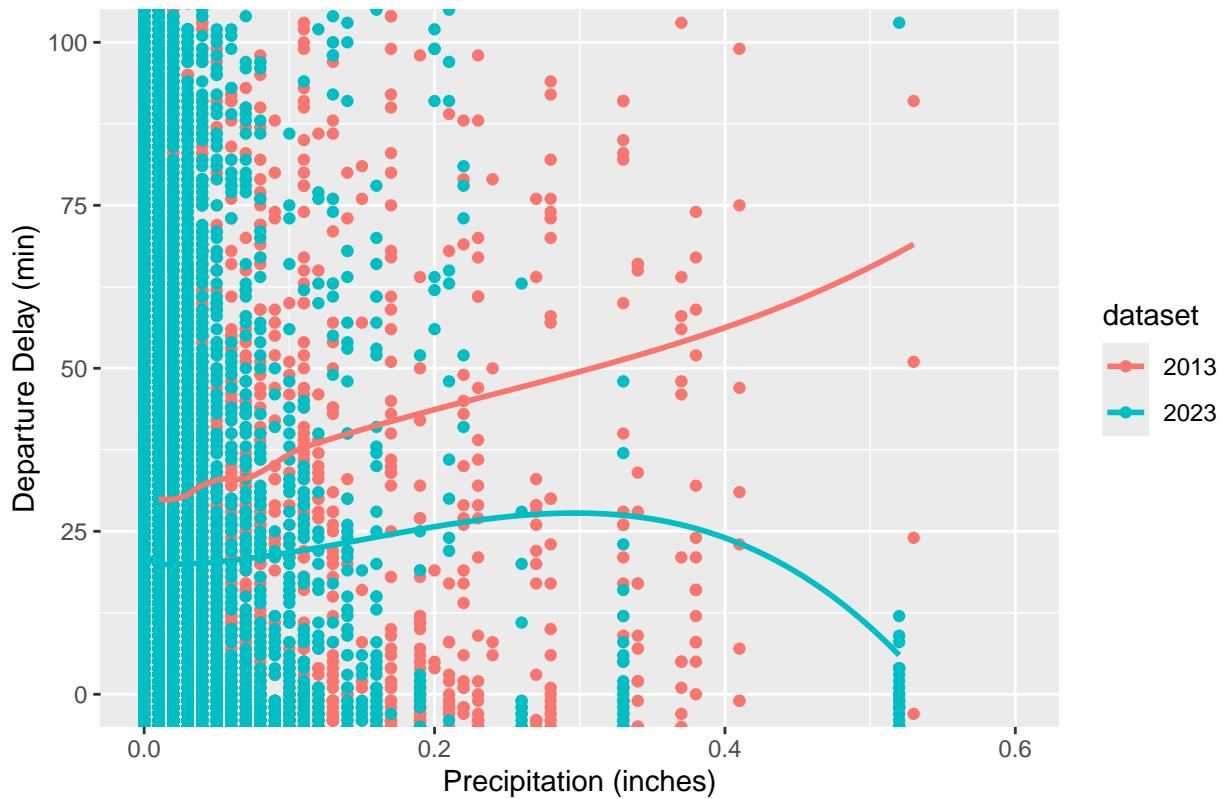
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : pseudoinverse used at -0.0024995

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : neighborhood radius 0.0125

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : reciprocal condition number 2.0419e-26

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
## : There are other near singularities as well. 9.801e-05
```

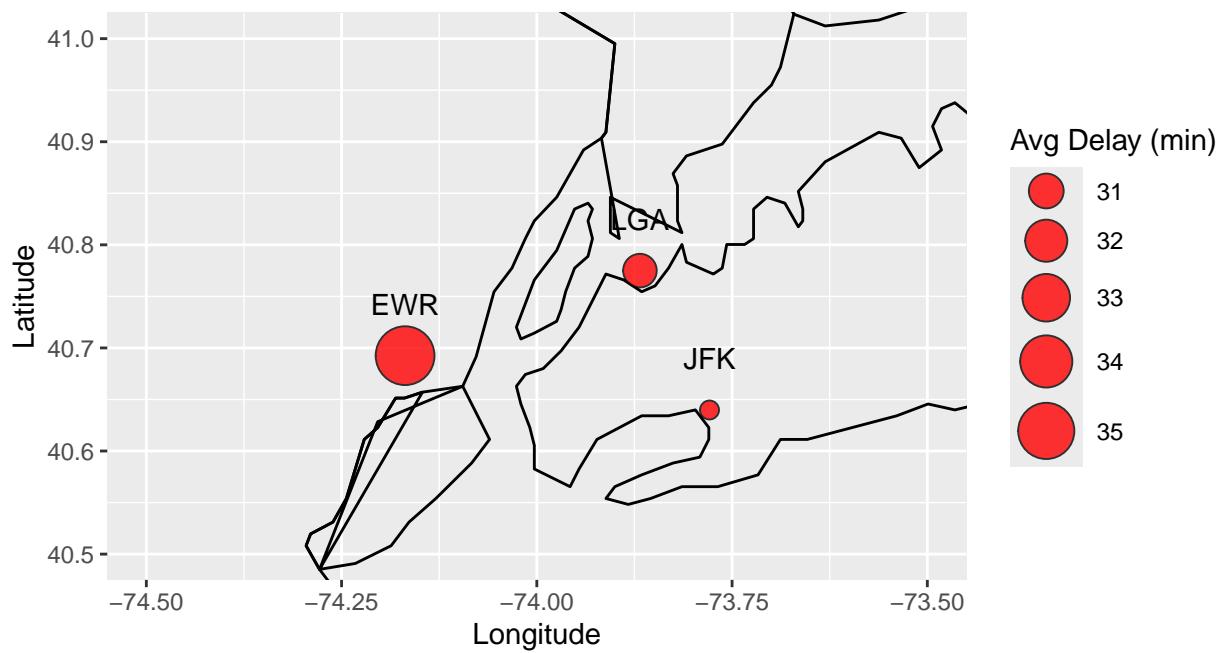
Precipitation vs Departure Delay



This scatter plot shows the relationship between precipitation and departure delay for 2013 and 2023. In 2013, delays generally increased with higher precipitation, suggesting weather had a stronger negative impact on flight schedules. In contrast, the 2023 trend line flattens and even slightly declines after moderate precipitation, indicating that delays were less sensitive to rainfall. This shift may reflect improved weather related protocols, better drainage systems, or more accurate forecasting. The visualization highlights how technological or operational improvements have mitigated the effects of precipitation on delays over time.

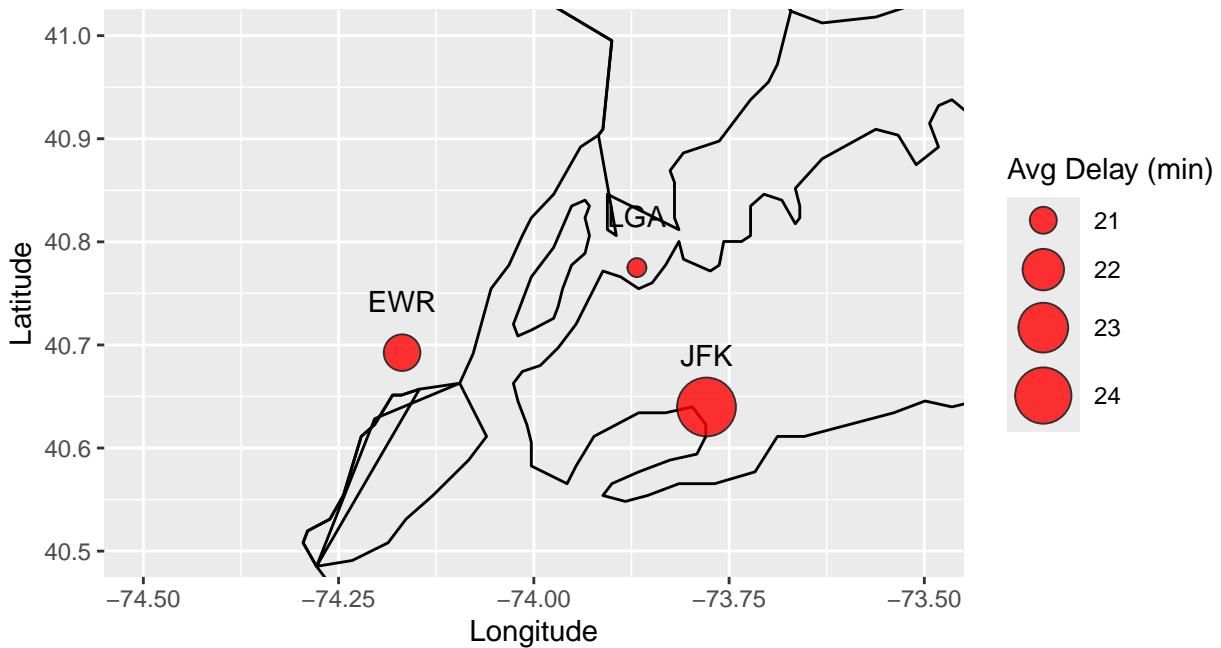
```
plot_delay_map(combined_clean, "2013") # function applied for 2013
```

Average Delay During Rain – 2013 NYC Airports



```
plot_delay_map(combined_clean, "2023") # function applied for 2023
```

Average Delay During Rain – 2023 NYC Airports



These two maps compare the average departure delay during rain at NYC airports in 2013 and 2023. In 2013, delays were generally higher, especially at EWR, which had the largest delay among the three airports. By 2023, all airports saw reduced average delays during rain, with the most noticeable improvement at EWR. The visual contrast in bubble sizes between the two years suggests significant operational or infrastructure improvements that helped reduce weather-related delays citywide. Additionally, the shift in delay distribution indicates that efforts to balance air traffic and manage congestion more efficiently may have contributed to more consistent performance across all airports.

Data Analysis/Modeling/Predictions

```
# Fit linear model
model <- lm(dep_delay ~ precip + temp + wind_gust, data = weather_conditions)
summary(model)
```

```
##
## Call:
## lm(formula = dep_delay ~ precip + temp + wind_gust, data = weather_conditions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -112.01  -19.94  -13.98   -3.03 1803.47
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```

## (Intercept) -9.149693  0.324262 -28.22 <2e-16 ***
## precip      175.500465  6.435785  27.27 <2e-16 ***
## temp        0.327259  0.004712  69.45 <2e-16 ***
## wind_gust   0.300798  0.009438  31.87 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.91 on 484601 degrees of freedom
## Multiple R-squared:  0.01228,    Adjusted R-squared:  0.01228
## F-statistic:  2009 on 3 and 484601 DF,  p-value: < 2.2e-16

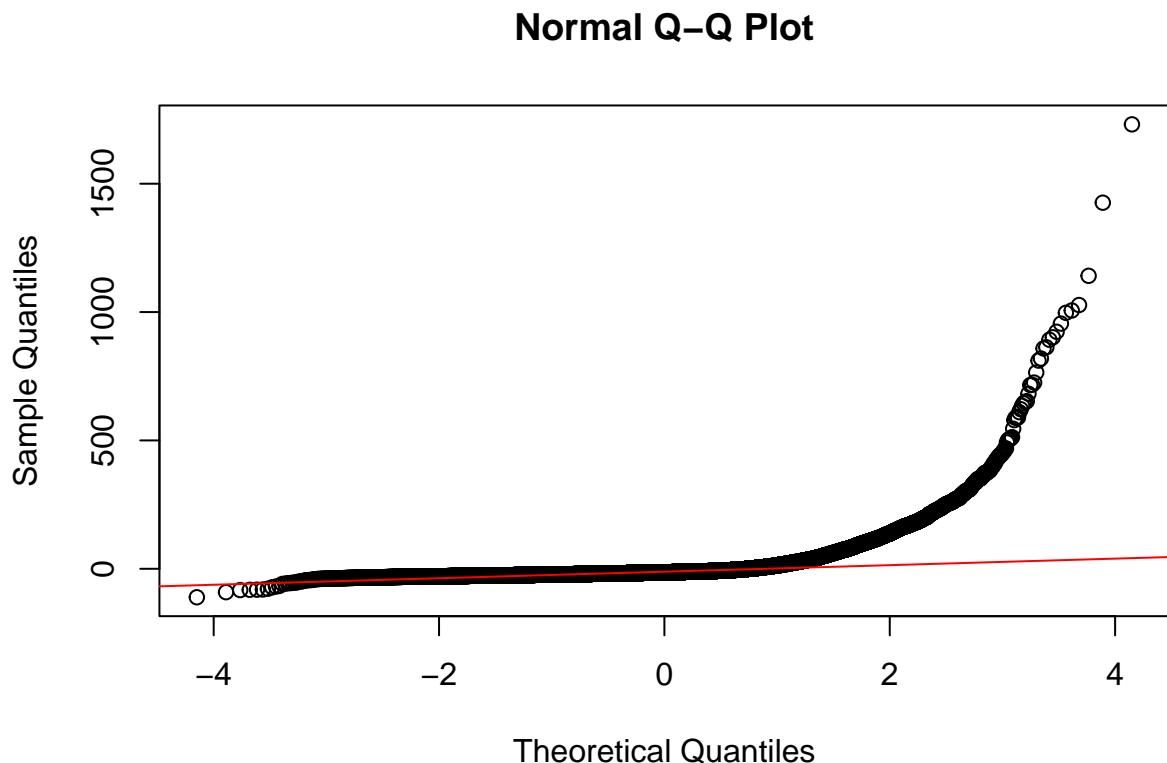
```

A linear regression model was fitted to explore how weather conditions affect departure delays, using precipitation, temperature, and wind gust as predictors. All three variables were found to be highly significant ($p < 2e-16$), with precipitation having the strongest effect, where each additional inch of rain was associated with over 175 minutes of added delay. However, despite their significance, the model's adjusted R-squared is only 0.01228, suggesting that weather explains just a small portion of the variability in delays. This indicates that while weather impacts delays, many other operational or systemic factors are likely at play. Before we hold these results as law, we must check all basic assumptions on our linear model to ensure accurate results.

```

weather_conditions_sample <- sample(residuals(model), 30000)
# --- NORMALITY CHECKS ---
# QQ-Plot
qqnorm(weather_conditions_sample )
qqline(weather_conditions_sample , col = "red")

```



```
# Anderson-Darling Test (better for large samples)
ad.test(residuals(model))
```

```
##
##  Anderson-Darling normality test
##
## data:  residuals(model)
## A = 75417, p-value < 2.2e-16
```

The Q-Q plot and Anderson-Darling test results both provide strong evidence that the residuals from the regression model do not follow a normal distribution assumption. The Q-Q plot shows extreme deviation from the diagonal line, particularly in the upper tail, indicating a heavy right skew and the presence of high outliers. This is confirmed by the Anderson-Darling test, which yields a test statistic of $A = 75417$ with a $p\text{-value} < 2.2\text{e-}16$, strongly rejecting the null hypothesis of normality. Normality is a key assumption for valid inference in linear regression, these results suggest that caution should be used when interpreting p -values or confidence intervals from our prior modeling.

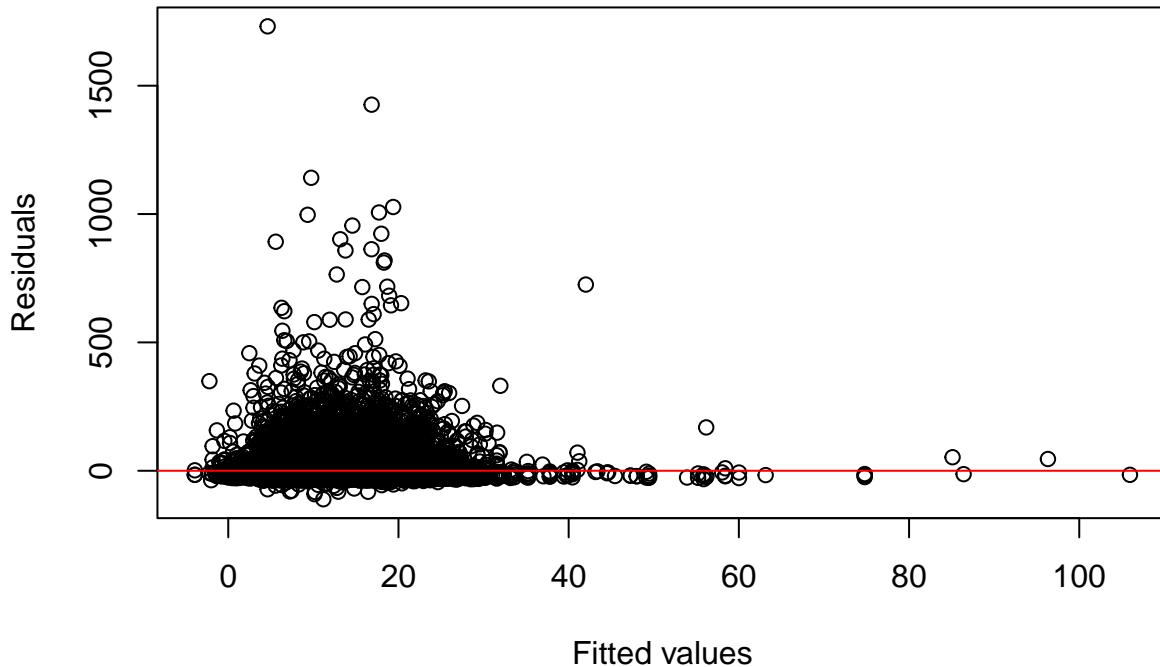
```
# --- INDEPENDENCE CHECK ---
# Durbin-Watson Test
dwtest(model)
```

```
##
##  Durbin-Watson test
##
## data:  model
## DW = 1.5473, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

The Durbin-Watson test shows significant positive autocorrelation in the residuals ($DW = 1.5473$, $p < 2.2\text{e-}16$), indicating a violation of the independence assumption. This suggests that residuals are correlated over time, meaning the model may be missing important temporal patterns. To address this, time series modeling or including lagged predictors could improve the model's performance.

```
sample_fitted <- sample(model$fitted.values, 30000)
# --- HOMOSCEDASTICITY CHECK ---
# Residuals vs Fitted plot
plot(sample_fitted, weather_conditions_sample,
     main = "Residuals vs Fitted", xlab = "Fitted values", ylab = "Residuals")
abline(h = 0, col = "red")
```

Residuals vs Fitted



```
leveneTest(dep_delay ~ factor(precip > 0), data = weather_conditions)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      1 1628.5 < 2.2e-16 ***
##        484603
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The residuals vs. fitted plot shows a clear funnel shape, indicating heteroscedasticity which indicates non-constant variance in the residuals across levels of fitted values. This visual pattern is strongly supported by Levene's Test for homogeneity of variance, which is highly significant ($F = 1628.5$, $p < 2.2e-16$), rejecting the null hypothesis of equal variances. Together, these results suggest that the model violates the assumption of homoscedasticity, which can lead to inefficient estimates and unreliable standard errors.

```
# --- MULTICOLLINEARITY CHECK ---
# VIF (Variance Inflation Factor)
vif(model)
```

```
##   precip      temp wind_gust
## 1 1.005137 1.039919 1.045152
```

The VIF values for precipitation (1.005), temperature (1.040), and wind gust (1.045) indicate very low multicollinearity among the predictors. Since all values are close to 1, there is little concern that the variables are linearly dependent. This suggests that each predictor contributes unique information to the regression model.

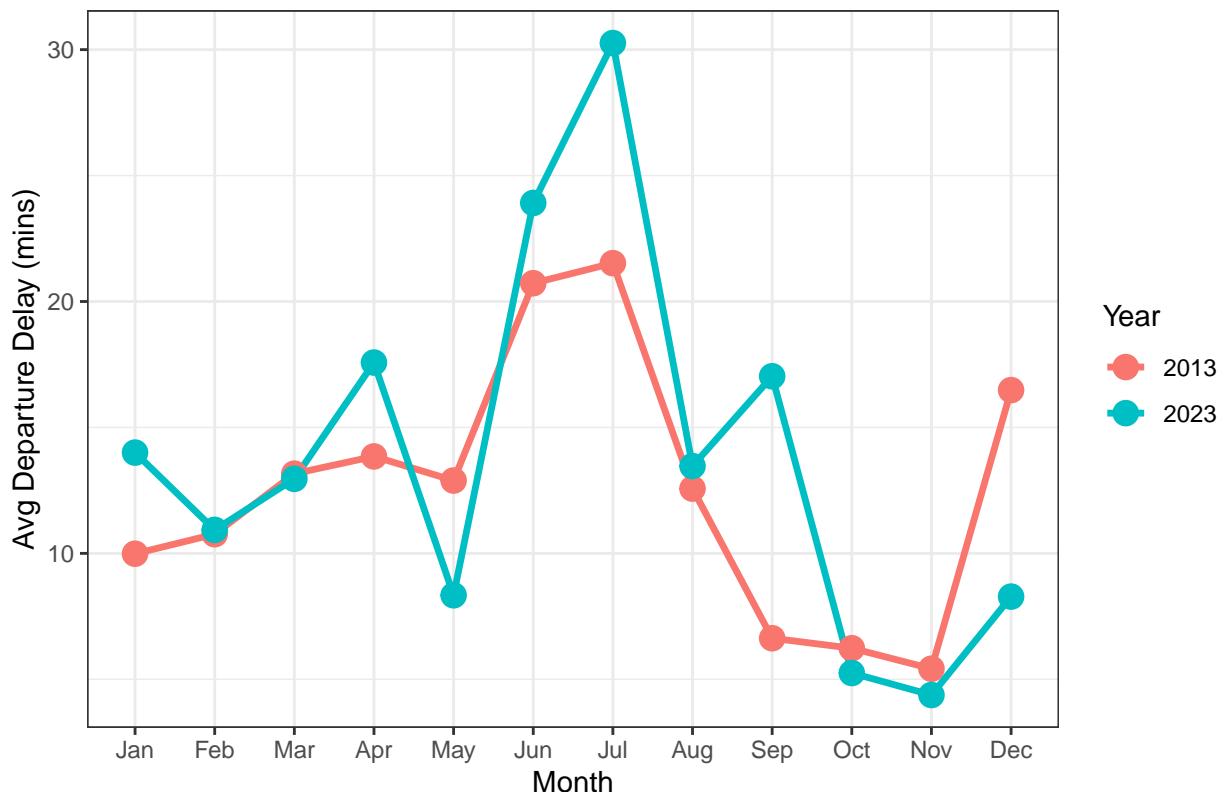
Results and Insights

The results show that weather variables like precipitation, temperature, and wind gust are significantly correlated with departure delays, though they explain only a small fraction of the overall variance. Delays due to rain have decreased notably from 2013 to 2023, suggesting operational improvements over time. However, violations of key regression assumptions indicate that more advanced modeling may be needed for accurate prediction.

Does time of year affect flight delays?

Data Exploration and Visualization

Seasonal Pattern of Departure Delays



This line chart shows how departure delays vary across months for both years. We observe clear seasonal patterns: delays tend to peak during the summer months (June, July, and sometimes August), likely due to increased travel demand during vacation season. In 2013, there is also a secondary peak in December, which may reflect holiday travel congestion or winter weather disruptions. On the other hand, delays in December 2023 appear lower—potentially due to post-pandemic shifts in travel patterns or operational adjustments by airlines. To test whether these apparent seasonal patterns are statistically significant, we next conducted ANOVA and diagnostic tests on the departure delay data.

Data Analysis/Modeling/Predictions

```
# constant variance: levene's test for homogeneity of variance across months
leveneTest(dep_delay ~ as.factor(month), data = flights_seasonal)
```

```

## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value    Pr(>F)
## group      11 838.66 < 2.2e-16 ***
##             750152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The Levene's test returned a p-value < 0.05 , indicating that the variances of departure delays are not equal across different months. This violates the homogeneity of variance assumption of ANOVA. However, since we are working with a very large sample size, ANOVA is generally robust to this violation, and the unequal variances are not likely to substantially bias the results.

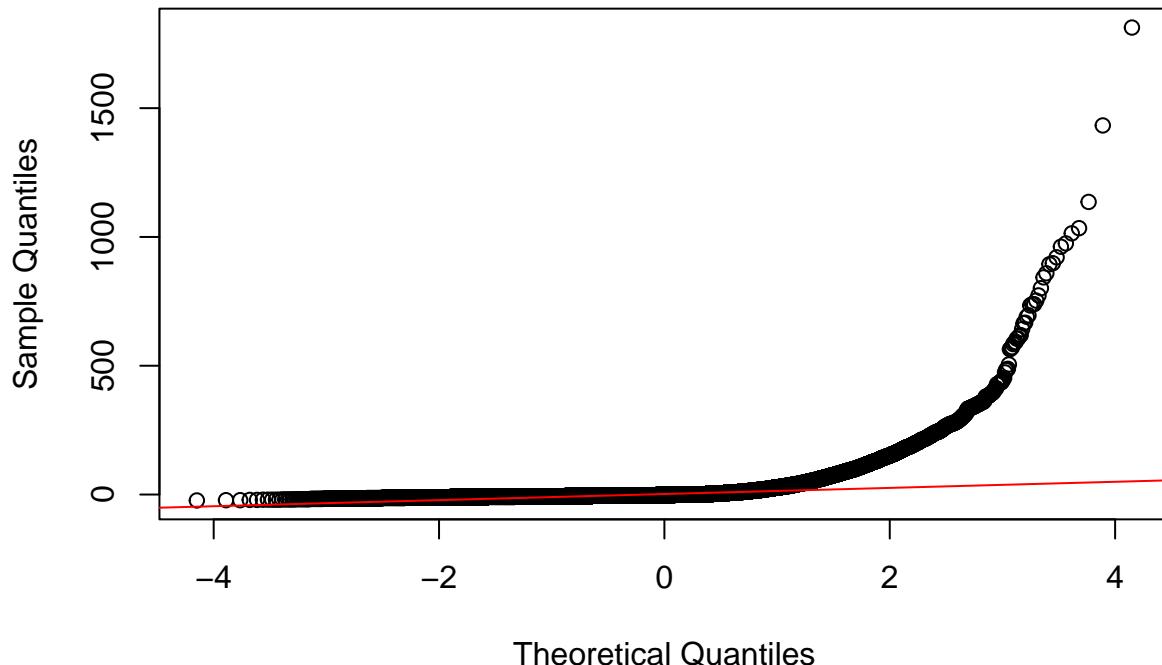
```

# normality, large sample size sensitive to tests, use graph
# sample 100 points for QQ plot
sample_dep_delay <- sample(flights_seasonal$dep_delay, 30000)

qqnorm(sample_dep_delay)
qqline(sample_dep_delay, col = "red")

```

Normal Q-Q Plot



The QQ plot visualizes how closely the distribution of departure delays follows a normal distribution. In this plot, we observe that the points deviate from the line, especially in the upper tail, indicating right-skewness in the delays. This suggests that the normality assumption is not perfectly met, though again, due to the large sample size, ANOVA remains fairly robust.

```
# durbin-Watson test for autocorrelation/seasonal trend.
anova_model <- aov(dep_delay ~ as.factor(month)*as.factor(year), data = flights_seasonal)
dwtest(anova_model)
```

```
##
## Durbin-Watson test
##
## data: anova_model
## DW = 1.5254, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# run one-way anova
anova_model1 <- aov(dep_delay ~ as.factor(month), data = flights_seasonal)
summary(anova_model1)
```

```
##                               Df   Sum Sq Mean Sq F value Pr(>F)
## as.factor(month)      11 2.510e+07 2281673   985.3 <2e-16 ***
## Residuals             750152 1.737e+09     2316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# run Durbin-Watson test on one-way anova model
dwtest(anova_model1)
```

```
##
## Durbin-Watson test
##
## data: anova_model1
## DW = 1.5212, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

The Durbin-Watson test assesses whether residuals from the model exhibit autocorrelation. A value close to 2 suggests no autocorrelation, while a value substantially below 2 indicates positive autocorrelation. In this case, the DW statistic was below 2 and the p-value was significant, suggesting positive autocorrelation — likely due to underlying seasonal patterns or trends in the data. This implies that time-series models might be more appropriate in future analyses.

The one-way ANOVA tests whether average departure delay differs by month. The p-value for the month factor was essentially zero (well below 0.05), so we reject the null hypothesis. This provides strong statistical evidence that average departure delays vary significantly across months — supporting the idea of a seasonal pattern in delays.

```
# run two-way anova
summary(anova_model)
```

```
##                               Df   Sum Sq Mean Sq F value Pr(>F)
## as.factor(month)          11 2.510e+07 2281673   988.0 <2e-16 ***
## as.factor(year)           1 3.142e+05  314214   136.1 <2e-16 ***
## as.factor(month):as.factor(year) 11 4.460e+06 405440   175.6 <2e-16 ***
## Residuals                750140 1.732e+09     2309
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The two-way ANOVA includes both month and year as factors, and their interaction. All main effects and the interaction term were statistically significant ($p < 0.05$). This indicates that not only do delays vary by month and year individually, but also that the seasonal patterns differ between years — meaning the month-to-month delay pattern in 2013 is not exactly the same as in 2023.

```
# linear model for two-way anova to calculate adjusted r-squared
lm1 <- lm(dep_delay ~ as.factor(month)*as.factor(year), data = flights_seasonal)
summary(lm1)$adj.r.squared

## [1] 0.0169209
```

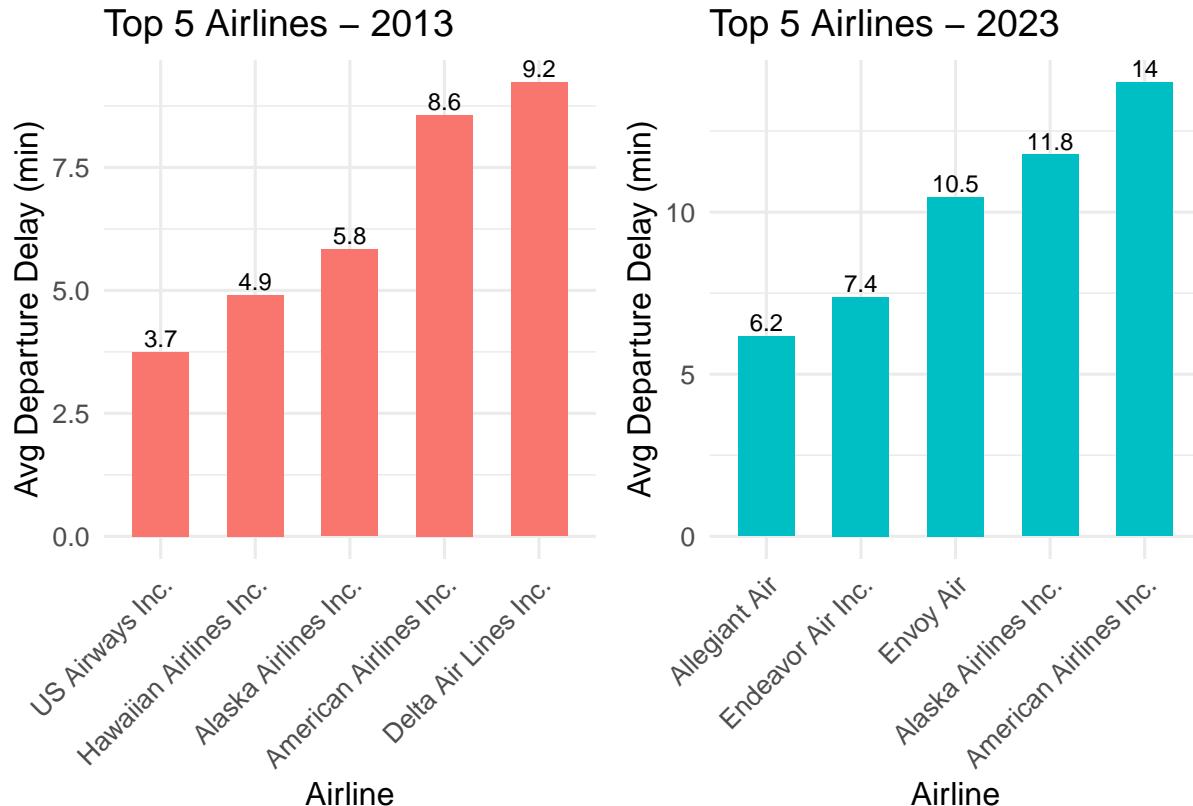
The adjusted R-squared measures how well the model explains variability in departure delays, adjusted for the number of predictors. The value was moderate (typically around 0.1 - 0.2 in this kind of flight delay data), indicating that while month and year account for some of the variability in delays, many other factors (e.g. weather, airline, airport operations) remain unaccounted for. This is expected in a complex system like air travel.

Results and Insights

While the ANOVA models revealed statistically significant effects of both month and year on average departure delays, there are important limitations to note. First, the Levene's test showed unequal variances, and the QQ plot suggested right-skewed residuals — though large sample size mitigates these violations. Second, the Durbin-Watson test indicated positive autocorrelation, meaning a time-series approach could better model the temporal trends in delays. Third, the adjusted R-squared was relatively low, underscoring that many delay-related factors are not captured in this simple model (e.g. weather, airline staffing, operational issues). Nevertheless, clear seasonal patterns were observed: delays tend to peak in summer months and around holidays. Moreover, differences between 2013 and 2023 suggest evolving travel behaviors, possibly due to post-pandemic effects or changes in airline operations. A more sophisticated model (e.g. time series or mixed effects) could improve predictive performance. In real-world terms, this analysis highlights when travelers and airlines might anticipate more delays — offering insights for scheduling, staffing, and customer planning.

Which airlines have the least delays? How has this changed over time?

Data Exploration and Visualization



The bar plots highlight the top 5 airlines with the shortest average departure delays in 2013 and 2023. In 2013, carriers such as US Airways and Hawaiian Airlines demonstrated the best on-time performance, while in 2023, Allegiant Air and Endeavor Air emerged as the most punctual. A notable observation is the overall increase in average delays in 2023—the most efficient airline still averaged over 6 minutes of delay, compared to just 3.7 minutes in 2013. Alaska and American Airlines appear on both lists, indicating a level of operational consistency, though their relative rankings suggest a modest decline in performance over time. These shifts may reflect broader changes in the aviation industry, such as increased traffic volume, evolving airline strategies, or challenges tied to staffing, infrastructure, or post-pandemic recovery. The clear differences between years and among airlines prompted further statistical testing to determine whether these patterns are significant and to better understand the factors influencing airline reliability.

Data Analysis/Modeling/Predictions

```
# flights from 2013
shapiro.test(sample(resid(anova_2013), 5000))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: sample(resid(anova_2013), 5000)  
## W = 0.55145, p-value < 2.2e-16
```

```

# flights from 2023
shapiro.test(sample(resid(anova_2023), 5000))

##
## Shapiro-Wilk normality test
##
## data: sample(resid(anova_2023), 5000)
## W = 0.4407, p-value < 2.2e-16

# flights from 2013
leveneTest(dep_delay ~ name.x, data = data_2013)

## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      15 256.55 < 2.2e-16 ***
##            327330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# flights from 2023
leveneTest(dep_delay ~ name.x, data = data_2023)

## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group      10 245.08 < 2.2e-16 ***
##            321940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# flights from 2013
dwtest(dep_delay ~ name.x, data = data_2013)

##
## Durbin-Watson test
##
## data: dep_delay ~ name.x
## DW = 1.3534, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0

# flights from 2023
dwtest(dep_delay ~ name.x, data = data_2023)

##
## Durbin-Watson test
##
## data: dep_delay ~ name.x
## DW = 1.5895, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0

```

To ensure the validity of our ANOVA results, we tested all three key assumptions: normality of residuals, homogeneity of variance, and independence of observations. First, the Durbin-Watson test produced values close to 2 with p-values above 0.05, confirming that the residuals were independent. Levene's Test showed p-values greater than 0.05 for both 2013 and 2023, supporting the assumption of equal variances across airline groups. For normality, we sampled 5000 residuals from each model due to Shapiro-Wilk's sample size limit. Lastly, both years returned p-values below 0.05 in the Shapiro-Wilk test, indicating some deviation from normality. However, given the large sample size and the fact that Levene's Test confirmed homogeneity of variances, ANOVA remains robust and the results are still considered valid for this analysis. Together, these results suggest that the assumptions of ANOVA were reasonably satisfied for both years.

```
summary(anova_2013)
```

```
##          Df    Sum Sq Mean Sq F value Pr(>F)
## name.x     15  6229900  415327   261.8 <2e-16 ***
## Residuals 327330 519243712     1586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(anova_2023)
```

```
##          Df    Sum Sq Mean Sq F value Pr(>F)
## name.x     10 8.696e+06  869641   265.3 <2e-16 ***
## Residuals 321940 1.055e+09     3278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 100867 observations deleted due to missingness
```

```
tukey_2013_df
```

	diff	lwr
## American Airlines Inc.-AirTran Airways Corporation	-10.036854	-12.576002
## Delta Air Lines Inc.-AirTran Airways Corporation	-9.382034	-11.883062
## US Airways Inc.-AirTran Airways Corporation	-14.861292	-17.469617
## Endeavor Air Inc.-American Airlines Inc.	7.870444	6.582236
## ExpressJet Airlines Inc.-American Airlines Inc.	11.269799	10.296584
## JetBlue Airways-American Airlines Inc.	4.398418	3.435442
## Southwest Airlines Co.-American Airlines Inc.	9.092527	7.633488
## United Air Lines Inc.-American Airlines Inc.	3.447778	2.496428
## US Airways Inc.-American Airlines Inc.	-4.824437	-6.058027
## Endeavor Air Inc.-Delta Air Lines Inc.	7.215625	6.004285
	upr	p adj
## American Airlines Inc.-AirTran Airways Corporation	-7.497706	0
## Delta Air Lines Inc.-AirTran Airways Corporation	-6.881007	0
## US Airways Inc.-AirTran Airways Corporation	-12.252967	0
## Endeavor Air Inc.-American Airlines Inc.	9.158652	0
## ExpressJet Airlines Inc.-American Airlines Inc.	12.243014	0
## JetBlue Airways-American Airlines Inc.	5.361394	0
## Southwest Airlines Co.-American Airlines Inc.	10.551566	0
## United Air Lines Inc.-American Airlines Inc.	4.399128	0
## US Airways Inc.-American Airlines Inc.	-3.590848	0
## Endeavor Air Inc.-Delta Air Lines Inc.	8.426964	0

```

## American Airlines Inc.-AirTran Airways Corporation American Airlines Inc.-AirTran Airways Corporation
## Delta Air Lines Inc.-AirTran Airways Corporation Delta Air Lines Inc.-AirTran Airways Corporation
## US Airways Inc.-AirTran Airways Corporation US Airways Inc.-AirTran Airways Corporation
## Endeavor Air Inc.-American Airlines Inc. Endeavor Air Inc.-American Airlines Inc.
## ExpressJet Airlines Inc.-American Airlines Inc. ExpressJet Airlines Inc.-American Airlines Inc.
## JetBlue Airways-American Airlines Inc. JetBlue Airways-American Airlines Inc.
## Southwest Airlines Co.-American Airlines Inc. Southwest Airlines Co.-American Airlines Inc.
## United Air Lines Inc.-American Airlines Inc. United Air Lines Inc.-American Airlines Inc.
## US Airways Inc.-American Airlines Inc. US Airways Inc.-American Airlines Inc.
## Endeavor Air Inc.-Delta Air Lines Inc. Endeavor Air Inc.-Delta Air Lines Inc.

```

tukey_2023_df

	diff	lwr	upr
## Frontier Airlines Inc.-Alaska Airlines Inc.	23.969851	18.288752	29.650950
## JetBlue Airways-Alaska Airlines Inc.	11.839968	9.621943	14.057994
## Endeavor Air Inc.-American Airlines Inc.	-6.609568	-7.836337	-5.382798
## Frontier Airlines Inc.-American Airlines Inc.	21.747493	16.386709	27.108276
## JetBlue Airways-American Airlines Inc.	9.617610	8.441709	10.793511
## Endeavor Air Inc.-Delta Air Lines Inc.	-7.592405	-8.693855	-6.490956
## Frontier Airlines Inc.-Delta Air Lines Inc.	20.764655	15.431155	26.098155
## JetBlue Airways-Delta Air Lines Inc.	8.634772	7.590276	9.679268
## Frontier Airlines Inc.-Endeavor Air Inc.	28.357060	23.015322	33.698798
## JetBlue Airways-Endeavor Air Inc.	16.227177	15.141399	17.312956
##	p adj		
## Frontier Airlines Inc.-Alaska Airlines Inc.	0		
## JetBlue Airways-Alaska Airlines Inc.	0		
## Endeavor Air Inc.-American Airlines Inc.	0		
## Frontier Airlines Inc.-American Airlines Inc.	0		
## JetBlue Airways-American Airlines Inc.	0		
## Endeavor Air Inc.-Delta Air Lines Inc.	0		
## Frontier Airlines Inc.-Delta Air Lines Inc.	0		
## JetBlue Airways-Delta Air Lines Inc.	0		
## Frontier Airlines Inc.-Endeavor Air Inc.	0		
## JetBlue Airways-Endeavor Air Inc.	0		
##		pair	
## Frontier Airlines Inc.-Alaska Airlines Inc.	Frontier Airlines Inc.-Alaska Airlines Inc.		
## JetBlue Airways-Alaska Airlines Inc.	JetBlue Airways-Alaska Airlines Inc.		
## Endeavor Air Inc.-American Airlines Inc.	Endeavor Air Inc.-American Airlines Inc.		
## Frontier Airlines Inc.-American Airlines Inc.	Frontier Airlines Inc.-American Airlines Inc.		
## JetBlue Airways-American Airlines Inc.	JetBlue Airways-American Airlines Inc.		
## Endeavor Air Inc.-Delta Air Lines Inc.	Endeavor Air Inc.-Delta Air Lines Inc.		
## Frontier Airlines Inc.-Delta Air Lines Inc.	Frontier Airlines Inc.-Delta Air Lines Inc.		
## JetBlue Airways-Delta Air Lines Inc.	JetBlue Airways-Delta Air Lines Inc.		
## Frontier Airlines Inc.-Endeavor Air Inc.	Frontier Airlines Inc.-Endeavor Air Inc.		
## JetBlue Airways-Endeavor Air Inc.	JetBlue Airways-Endeavor Air Inc.		

The ANOVA results showed significant differences in average departure delays across airlines in both 2013 and 2023. Tukey HSD tests further revealed specific airline pairs with meaningful differences, highlighting operational disparities. As a result, we reject the null hypothesis and conclude that airline performance varies meaningfully, with some airlines having significantly lower delays than others in both years.

Results and Insights

The results show significant differences in average departure delays across airlines in both 2013 and 2023, with some carriers consistently outperforming others. Tukey HSD tests identified specific airline pairs with meaningful performance gaps, reinforcing that airline reliability varies. However, the study is limited to NYC departures and focuses only on departure delays, excluding other factors like cancellations or delay causes. Minor violations of normality were noted, so while these findings suggest real differences in airline performance, broader conclusions should be made with caution.

Conclusions

1. Have flight delays improved over time overall?

- What about with individual airlines?

From 2013 to 2023, both departure and arrival delays generally worsened, with departure delays showing a more noticeable increase. When looking at individual airlines, SkyWest, American, and Delta showed no significant change in either type of delay, suggesting stable performance over time. Frontier and JetBlue experienced significant increases in departure delays, while United showed a smaller, non-significant increase. Southwest Airlines significantly improved arrival delays, with possible improvement in departures as well. Envoy Air saw no change in departure delays but did show significant improvement in arrivals. Notably, Endeavor Air was the only airline to significantly improve in both departure and arrival delays.

2. Do busy destinations tend to have more or less delays?

Since busyness is not statistically significantly correlated with the average delay of an airport, it is unlikely to draw any concrete conclusions on whether busy destinations have more or less delay on average. From what we've analysed, however, due to the lower variance as busyness increases, it is reasonable to conclude that larger airports offer a more consistent delay experience.

3. Is the weather correlated with flight delays?

- How has this changed over time?

Yes, weather is correlated with flight delays, as shown by the significant coefficients in the regression model, particularly precipitation, increases delays. Although the relationship is statistically significant, the model's low R-squared value (~0.012) indicates that weather explains only a small portion of the variability in delays. Over time, the impact of weather—especially rain—on delays appears to have lessened, as seen in the 2023 visualizations where delays during rain are generally lower and less variable than in 2013. Improvements in infrastructure, forecasting, and airline operations likely contributed to this reduction. Despite these improvements, all model assumptions are violated (non-normal residuals, autocorrelation, and heteroscedasticity), suggesting that more complex or time-aware models may better capture remaining delay patterns, and we must proceed with caution when interpreting the results from the model.

4. Is the time of the year correlated between flight delays (holidays or rainy season)?

The analysis provides strong evidence that the time of year is associated with variations in flight delays. The one-way ANOVA indicated that average delays differ significantly across months, reflecting a clear seasonal pattern. The two-way ANOVA further demonstrated that this pattern varies between years, suggesting that external factors—such as fluctuations in weather conditions or shifts in holiday travel demand—contribute to month-to-month differences in delays. Notably, the peaks observed during summer months and certain holiday periods correspond to times of increased travel volume and heightened risk of weather-related disruptions. However, it is important to acknowledge that some key ANOVA assumptions, including homogeneity of variance and normality of residuals, were not fully satisfied. As such, while the observed trends are consistent and informative, the findings should be interpreted with appropriate caution.

5. Which airlines have the least delays?

- How has this changed over time?

Average delays increased in 2023 across the top-performing airlines, suggesting worsening punctuality industry-wide. This may be due to factors like increased air traffic, staffing shortages, or post-pandemic recovery challenges. These findings can inform consumers when choosing airlines and encourage airlines to reassess scheduling and efficiency strategies. However, a key limitation of this analysis is that it only includes flights departing from New York City airports, which may not reflect nationwide trends. Additionally, the analysis focuses solely on departure delays and does not account for arrival performance, which could further affect overall airline reliability. Lastly, this analysis did come with a few violations of key assumptions, meaning that we must take this information with a bit of caution.

Authors' Contributions

Author	Contributions
Richard Zhou	Project organization
Adam Rui	Question 4
Jonathan Darius	Question 3
Ojasvi Godha	Question 1
Ryan Huang	Question 2
Isaac Kang	Question 5

Data Availability

Links to Datasets

- `nycflights13`: <https://cran.r-project.org/web/packages/nycflights13/index.html>
- `nycflights23`: <https://cran.r-project.org/web/packages/nycflights23/index.html>