The Pandemic's Effects on the Airline Industry:

Summer 2019 vs Summer 2020

STAT 3350.001 Project

Group 1

Ved Nigam, Hiba Khan, Monique Picott

Nov 15, 2022

Table of Contents

1 - Introduction

For this project, we were assigned to the cohort focused on JHU Covid-19 Data, and we debated over many different potential focuses/motivations for the specifics of our analysis. One idea that stood out to all of our group members was running an analysis on how the pandemic affected air travel. We all have majorly been affected by COVID-19 on an individual basis, so it intrigued us to see how it may have affected a major industry.

Diving into further research/investigation, we noted that the COVID-19 pandemic brought a sudden halt to the ability for people to travel freely. Many airlines took heavy hits and had to completely re-establish their protocols for flights to accommodate for COVID regulations (ex. social distancing and cleaning protocols). Even at first glance, in 2020-present database records, the effects of the pandemic in airline efficiency show clearly. We the did a thorough run through of the potential datasets on the provided website and chose the one that best displayed how the pandemic affected flight corporations on a major scale, hoping to see how they were able to adapt and power through to avoid bankruptcy and return to pre-pandemic efficiency/revenue.

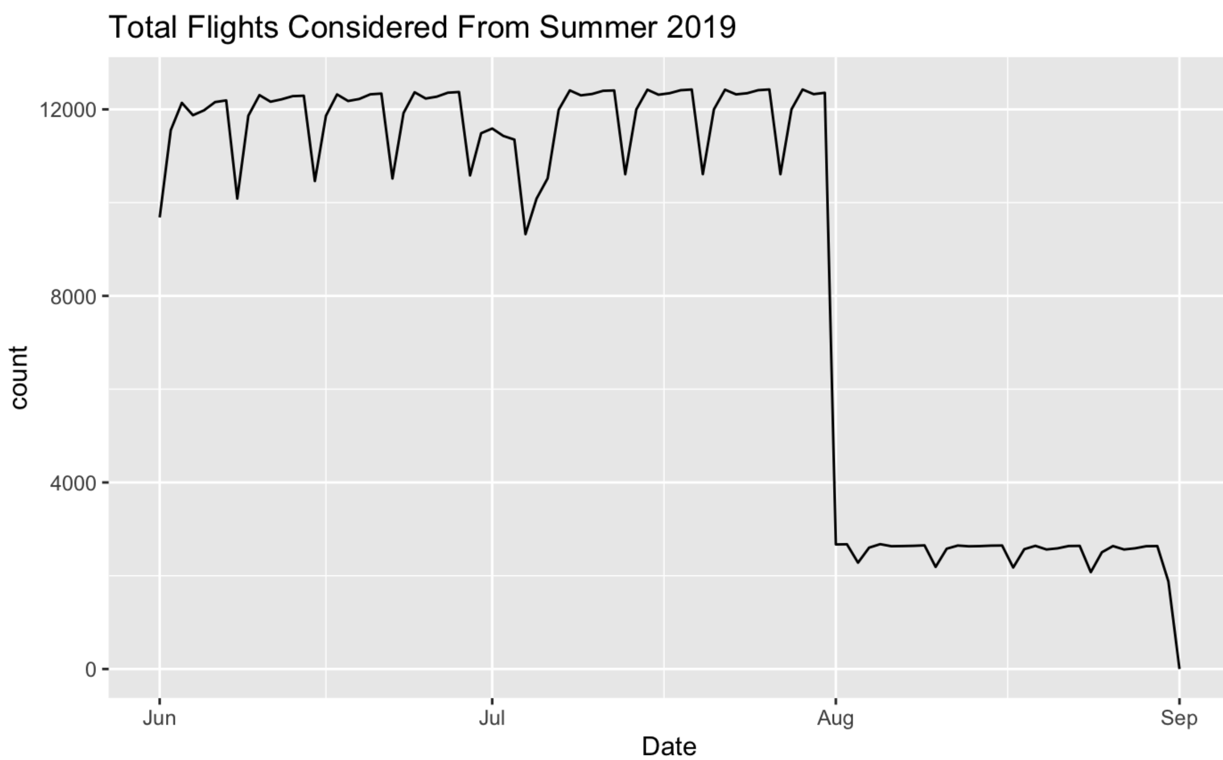
Our dataset, found on www.transtats.bts.gov, an official website run by the United States Department of Transportation, is a collection of data compiled by The Bureau of Transportation Statistics. Overall, it contains 2,348,374 observations and 22 variables. As our primary focus, we chose to focus on the summers of 2019 and 2020 to compare, as summers are statistically cited as the busiest months for air travel. We selected and subsetted the data for June, July, and August. The aviation data library contains 27 databases, and out of these, we have chosen to focus on the dataset Reporting Carrier

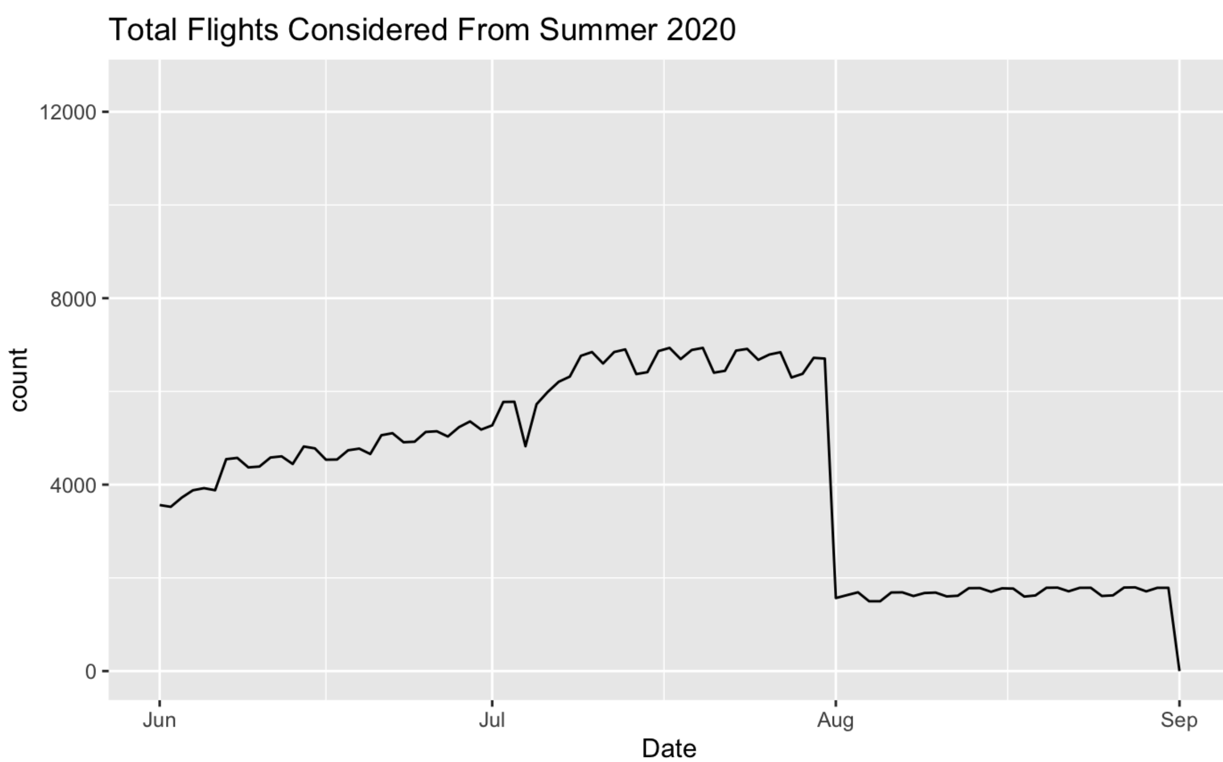
On-Time Performance Data in the state of Texas. We will be utilizing data from the years 2019 and 2020 which each contain 123,148 and 78,056 records, respectively. Total records that will be considered will be 201,204 records. Each dataset will contain the same information detailed through 21 variables. This dataset encompasses statistics on major airlines, their scheduled flights, delays, diversions, etc.

2 - Data Cleaning

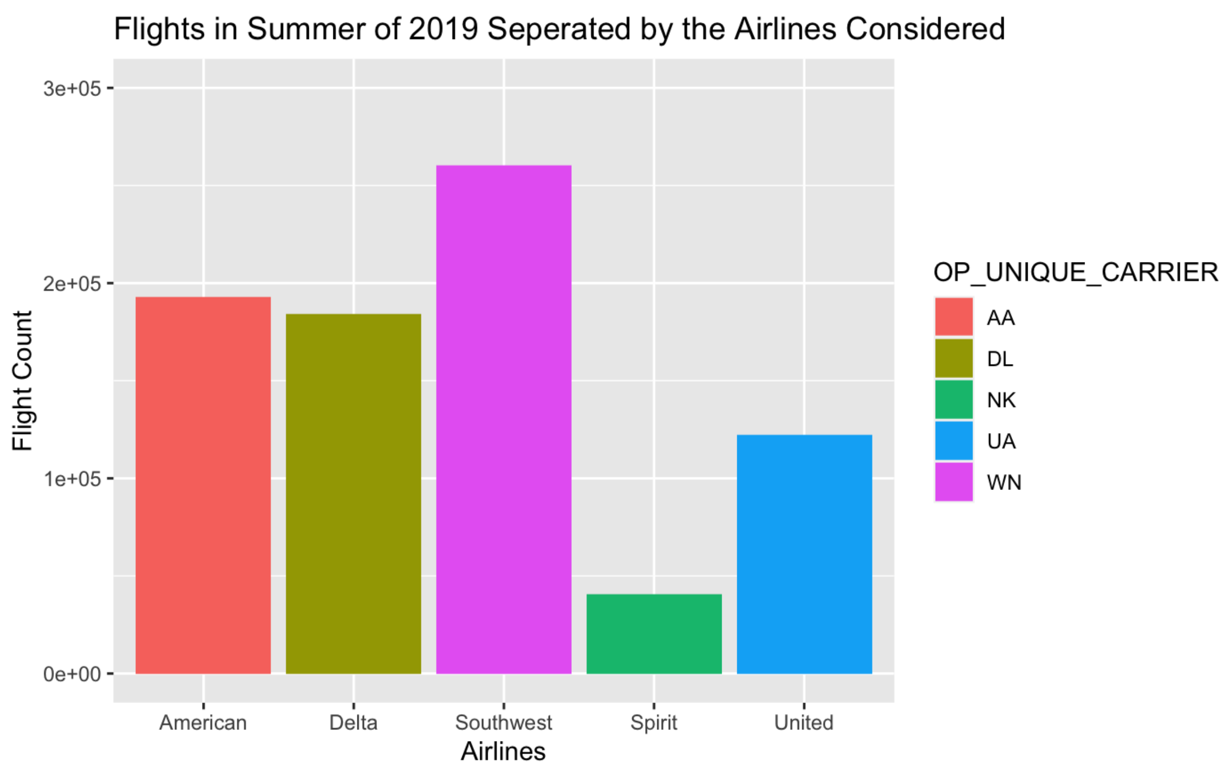
A significant aspect of having consistent and reliable test results is having clean data. The term ‘clean data’ refers to data that minimizes outliers, and is a dataset that focuses on the purpose of the research that the individual or group is conducting. For our purpose, we first picked specific time periods of data we wanted to work with (June-August of 2019 and 2020) and the specific airlines we wanted to work with (American, Delta, Southwest, Spirit, and United). After we sorted through our data, we began to subset it based on the variables we wanted to observe. An example of this would be subsetting the data into a dataset that contains only the data for the flights canceled. We also have subsets for different reasons of cancellation and for the departures/arrivals from some of the popular cities in the USA for domestic travel. For the entirety of the project, we kept the datasets for 2019 and 2020 separate because the primary focus of our research was to find differences/patterns/results between the two years due to the impact of COVID-19.

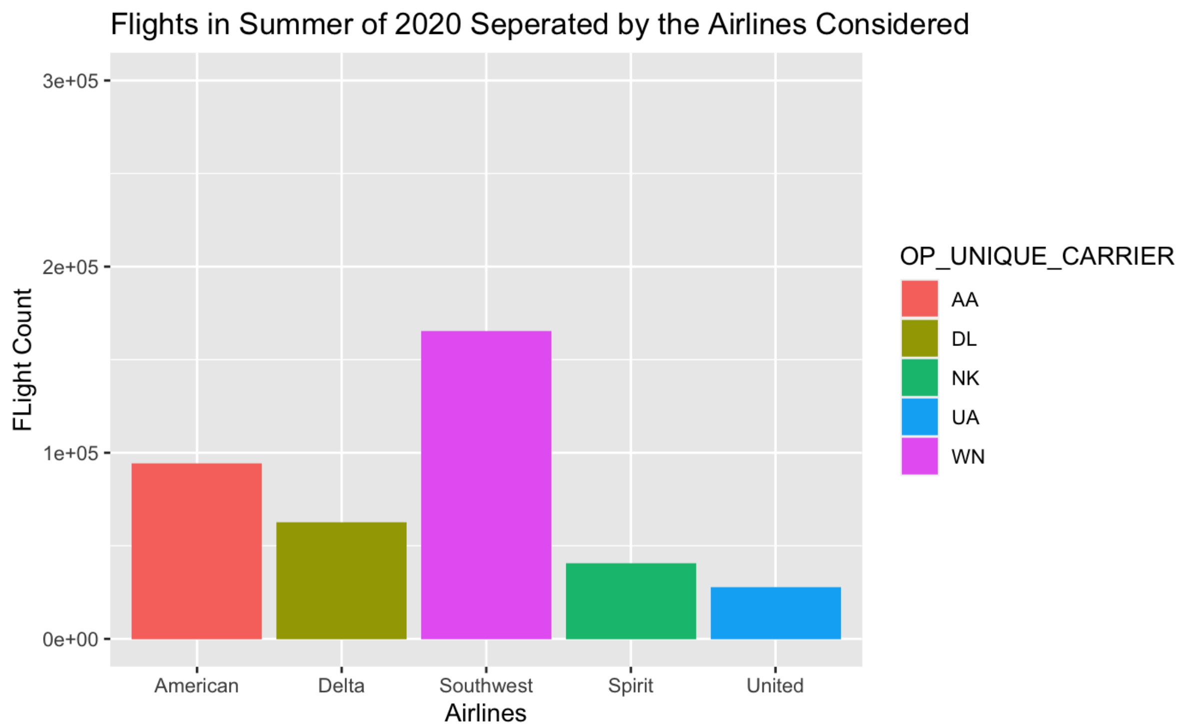
3 - Analysis: Comparison of Total Flights





To start off our analysis, we decided to take a look at the difference in the amount of flights scheduled each summer. We expected to see a significant drop in the amount of flights scheduled, as heavy travel restrictions were implemented not only in the US, but worldwide, as many areas became red zones for the virus and wouldn't allow travel in and out accordingly. Our dataset provided information for numerous airlines, but we decided it was best to analyze the top five major airlines in our dataset. After completing thorough research, we selected Delta Airlines, United Airlines, American Airlines, Southwest Airlines, and Spirit Airlines. For this particular graph, we thought the data would best be represented using a bar graph or a line graph, so we attempted to generate both.



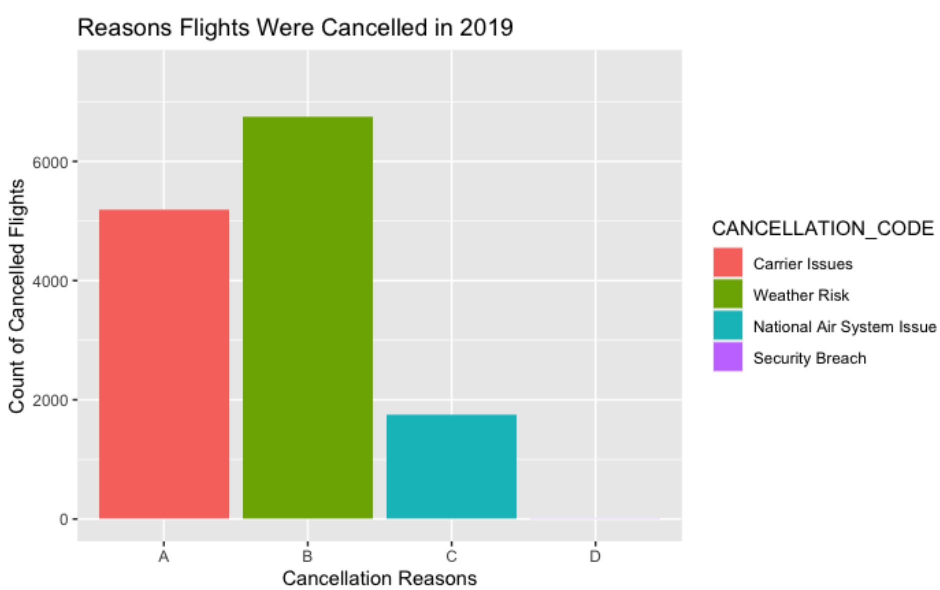


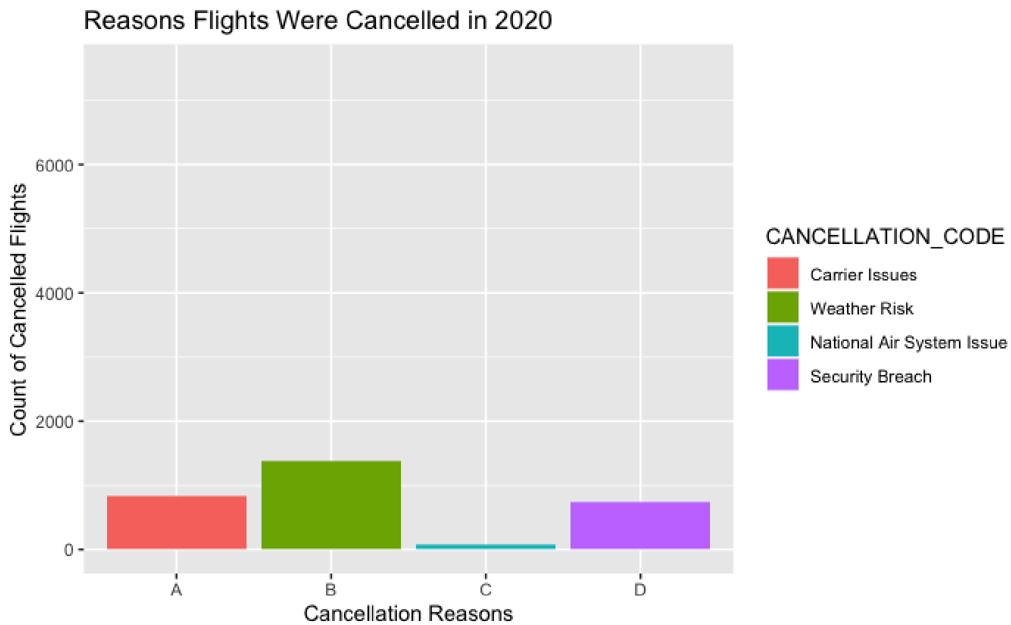
As can clearly be seen in the graphs, all the major airlines we chose to observe experienced a similar trend of decline with a significant drop in number of flights from summer 2019 compared to summer 2020. Delta Airlines saw a drop in 87,770 flights, which was a 47.66% decrease, United Airlines saw a drop in 85,147 flights which was a 69.58% decrease, American Airlines saw a drop in 73,412 flights which was a 38.09% decrease, Southwest Airlines saw a drop in 31,021 flights which was a 11.92% decrease, and lastly, Spirit Airlines saw a drop 4,226 flights which was a 10.41% decrease.

As can be seen by these percentages, United Airlines saw the greatest negative impact and Spirit Airlines experienced the smallest negative impact. Reportedly, United Airlines saw losses in profit upwards of 7 billion dollars, and although the percent decrease for Spirit Airlines was only 10.41 percent, reports still show that the corporation saw losses in the tens of millions of dollars in profit. Even with this heavy setback, thankfully, none of the airlines in our analysis had to file for bankruptcy, and now have returned to profit numbers similar to those of pre pandemic times.

When analyzing these graphs, it's important to take into account that passengers per flight decreased due to social distancing protocols; the United Nations cited an approximately 60% decrease in number of passengers on flights due to social distancing protocols and lack of voluntary travel, so regardless of the fact that some of these airlines were able to maintain the numbers of the flights they scheduled, incoming profit was still limited due to internal decreases on said flights. Additionally, it's important to account for the fact that airlines such as Delta, American, and United are frequently flown by individuals in the workplace. During the period when companies switched to remote work, travel for work essentially hit zero, which greatly contributed to these decreases.

4 - Analysis: Comparison of Reasons for Flight Cancellation



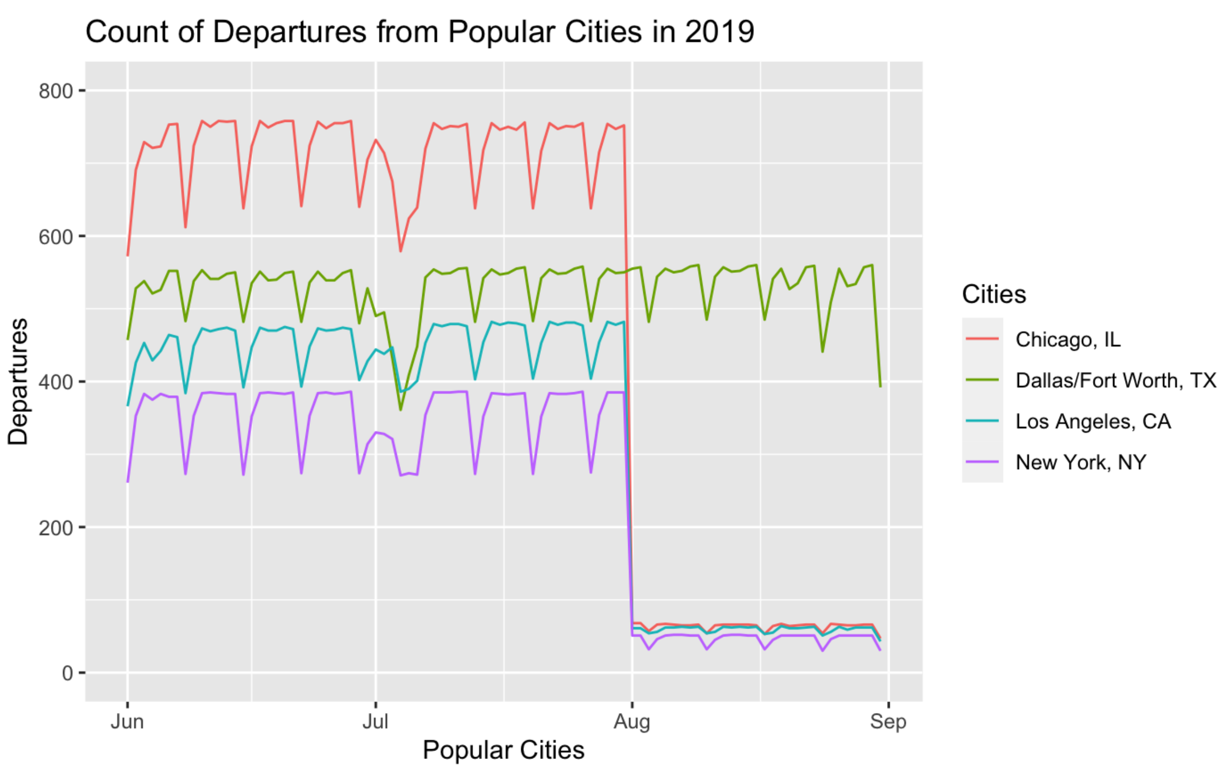


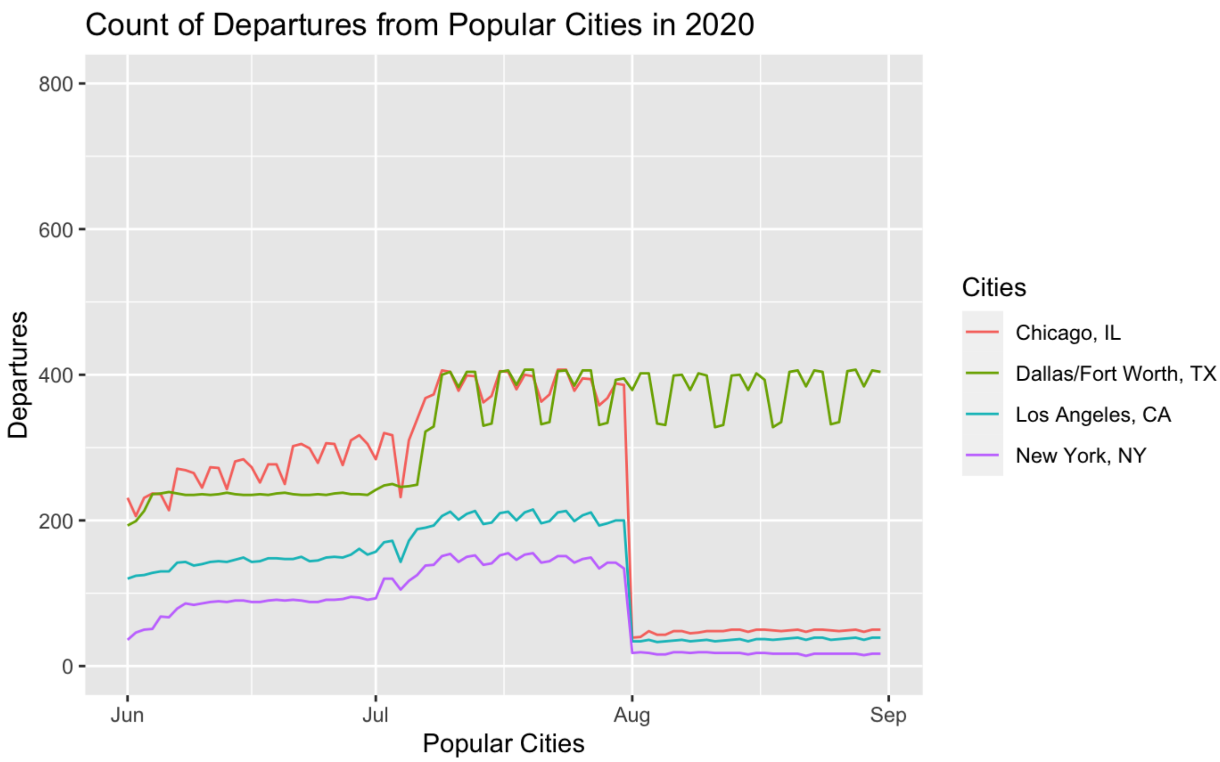
For our next analysis, we wanted to get a little bit deeper, and were interested in how the pandemic affected the cancellation of flights, specifically, the reasons behind flights getting canceled. We again opted to use a bar chart for our analysis. The dataset we chose categorizes the reasons for flight cancellations in four groups: carrier issues, which encompasses mechanical/technical issues with the plane, inclement weather, which encompasses dangerous conditions to fly in due to unexpected heavy weather, National Air System (NAS) issues, which encompass scheduling and administrative issues, and lastly security risks, which accounts for any flights canceled due to a breach of safety.

According to our graphs, carrier issues issues saw a drop in 3,926 cancellations, which was a 75.49% decrease, inclement weather saw a drop in 5,675 cancellations, which was a 84.09% decrease, NAS issues saw a drop in 1,658 cancellations, which was a 94.85% decrease, and security risks saw an increase in 1,407 cancellations, which was a 99.72% increase. Due to the fact we had less flights scheduled in the first place, it makes sense that we would have less flights canceled per category. However, we can observe that one category is not like the others! Security risks saw an almost one hundred percent increase.

According to the variable descriptions on our datasets website, biohazards and pathogen exposure are classified as security risks, so in times where Covid-19 had higher risk of transmission with new variants being identified and cities becoming redzones, flights would be canceled and classified as security risks. We had few expectations of what our data would indicate, but the results were quite straightforward and easy to analyze.

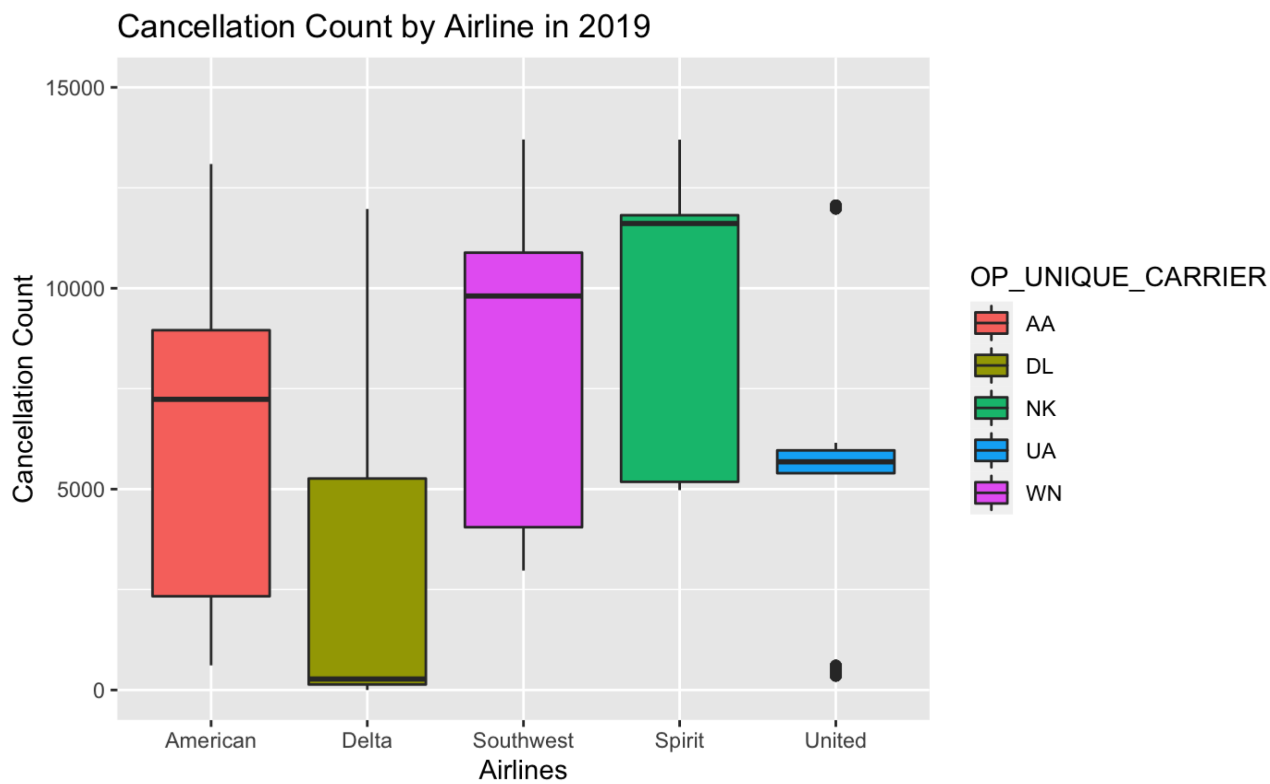
5 - Analysis: Comparison of Arrivals and Departures from Busiest Airports in the USA

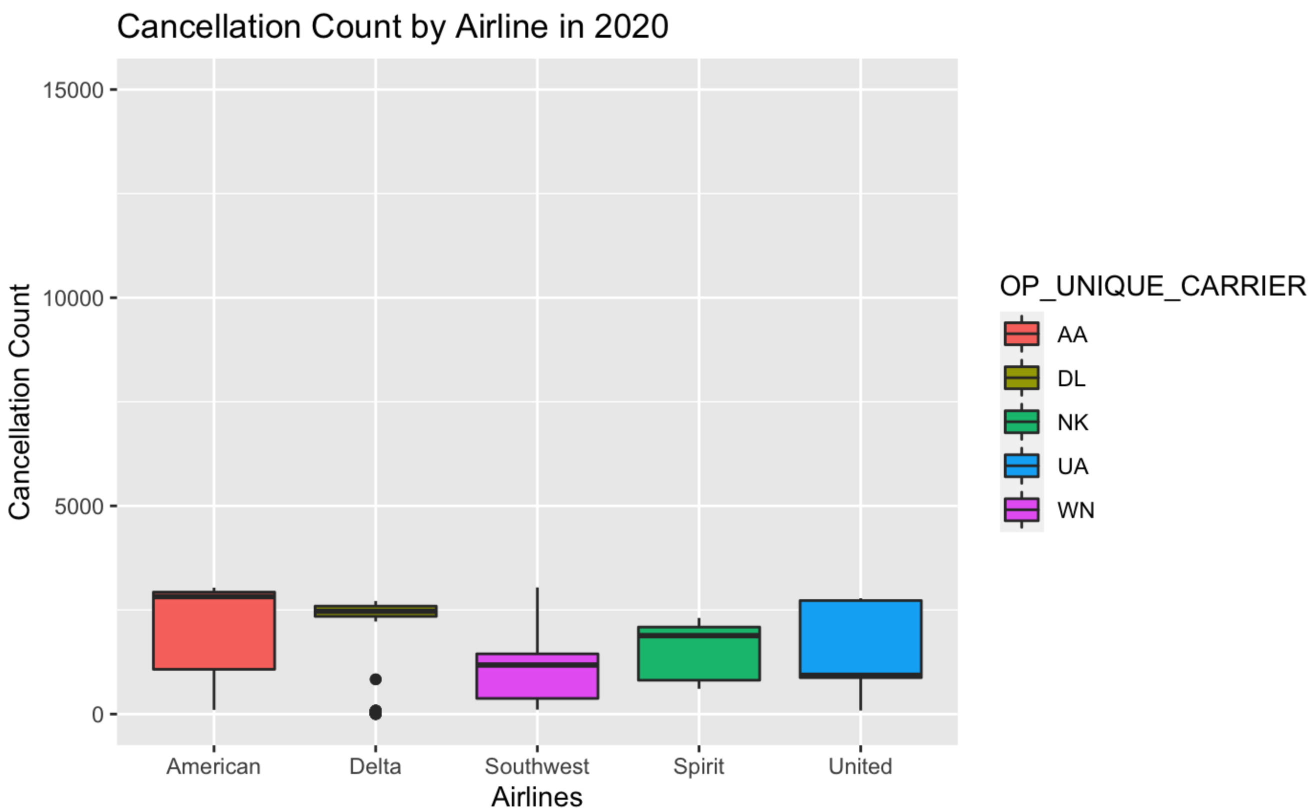




Next we decided to focus on departures and arrivals of popular/busier cities. Focusing on some of the more popular or busier cities in the US, such as Chicago, Dallas, Los Angeles, and New York, we observed a significant decrease in departures and arrivals from Summer 2019 to Summer 2020. Flight departures went down by 15,282, or a little over a 41% decline. This decrease happened as a direct effect of the pandemic. Many airlines offered fewer flights during 2020. This declining trend correlates with the restrictions and limitations many airlines began to experience. As observed in our chart it displays higher numbers in 2019 however in 2020 a significant drop can be observed. Fewer passengers traveled during this time possibly from fear of catching covid or spreading it. It is important to note that many of these cities implemented travel restrictions, some even had curfews and lockdowns. These restrictions and factors most likely played a role in the decline of departures.

6 - Analysis: Cancellation Comparison Between Airlines





For our last set of analysis, we thought it would be interesting to observe which summer had the greater percentage of flights canceled. To do this, we again chose the same 5 major airlines from before and did comparisons on the amount of flights canceled per airline each summer. At first, we opted to use pie charts, but in the end opted to use box plots to better display the data. We expected to see an increase in cancellations due to the pandemic, but once we drew up our data and generated our graphs, we were surprised by our findings. The number of cancellations decreased relatively proportionally to the decrease in the number of flights, so the actual percent of flights canceled did not significantly change for any of the airlines.

To understand our resulting graphs, we tried to account for the factors that would have created these results. Specifically, we have to consider the time range we selected. It’s likely that if we had run analysis on spring (March/April) we would have seen the expected major increase in flight cancellations, as that is when the pandemic’s effects began and by the time corporations reached summer, they would have adjusted scheduling accordingly.

For a deeper and more thorough analysis, there are many other factors we should take into account before drawing conclusions. We could account for quantified flight demand, and number of available employees/allowed cap on employees dues to social distancing protocols. Not only does this include flight attendants/cabin crew, but pilots and other technical flight staff. Factors such as these were not accounted for in our dataset, and to be able to account for these we would have to link more data from additional sets for a more accurate and broader analysis.

7 - Conclusion

Finally, from extensive research and analysis of our data, our team was able to observe the effects of COVID-19 substantial impact on some of America's major airlines. Upon further observation, we chose to examine the summer months of June, July, and August more closely. Summer months are ideal when researching because these months allow people to travel more. Children are out of school for summer break, families want to try to spend time, and the weather is more desirable. We found a significant drop in travel from summer 2019 to summer 2020. This drop corresponds with the travel restrictions and limitations that came as a direct response to COVID-19. Our team had to take into consideration several factors when examining our data. For one, the coronavirus did not have the same effect on every city in the United States of

America. Cities considered popular destinations for major airlines were now also considered to be outbreak centers for the coronavirus due to the large populations and attraction of tourism. With restrictions limiting the number of passengers on flights and demand decreasing, many airlines reduced the amount of flights offered.

Even though restrictions limited airlines, cancellations, surprisingly, did not see an increase but rather a decline. Cancellations usually occur because of weather or carrier issues; however, with covid, airlines began to reduce the number of flights because the demand was not as substantial. This reduction also helped to decrease the number of cancellations significantly.

Although our report had over 2 million observations and over 20 variables, it is essential to understand that we could only examine some factors closely. In our future work, we plan to analyze the overall years, including international airlines in our research, improve our approach and reanalyze some of the dead-end questions we came across.We are curious why Southwest airlines did not see a considerable decline like its competitors. We would also like to consider the weeks or days individually rather than looking at months as a whole. Some days are more popular for travel than others, and taking a closer look at other seasons, not just summer months. Doing this will lead to better data for comparison.

8 - Code

```
# Reading in data
nineteenJune_data <- read.csv("airTravel2019June.csv", header = TRUE)
twentyJune_data <- read.csv("airTravel2020June.csv", header = TRUE)
```

```
nineteenJuly_data <- read.csv("airTravel2019July.csv", header = TRUE)
twentyJuly_data <- read.csv("airTravel2020July.csv", header = TRUE)

nineteenAug_data <- read.csv("airTravel2019Aug.csv", header = TRUE)
twentyAug_data <- read.csv("airTravel2020Aug.csv", header = TRUE)

# Condensing data by year
nineteen_flights <-
  rbind(nineteenJune_data, nineteenJuly_data, nineteenAug_data)
twenty_flights <-
  rbind(twentyJune_data, twentyJuly_data, twentyAug_data)

# All data considered
total_flights <- rbind(nineteen_flights, twenty_flights)
total_flight_count <- nrow(total_flights)

# Count of all flights considered in the year
nineteen_flights_count <- nrow(nineteen_flights)
twenty_flights_count <- nrow(twenty_flights)

#total number of flights in summer of 2019 and 2020:
print(nineteen_flights_count + twenty_flights_count)

# Used to find the airline codes in the dataset
unique(nineteen_flights$OP_UNIQUE_CARRIER)

# Subsetting data based on airline and year
# Summer of 2019
```

```
Delta_nineteen_data <-
  subset(nineteen_flights, nineteen_flights$OP_UNIQUE_CARRIER == "DL")
Delta_flights_nineteen <- nrow(Delta_nineteen_data)

United_nineteen_data <-
  subset(nineteen_flights, nineteen_flights$OP_UNIQUE_CARRIER == "UA")
United_flights_nineteen <- nrow(United_nineteen_data)

American_nineteen_data <-
  subset(nineteen_flights, nineteen_flights$OP_UNIQUE_CARRIER == "AA")
American_flights_nineteen <- nrow(American_nineteen_data)

Southwest_nineteen_data <-
  subset(nineteen_flights, nineteen_flights$OP_UNIQUE_CARRIER == "WN")
Southwest_flights_nineteen <- nrow(Southwest_nineteen_data)

Spirit_nineteen_data <-
  subset(nineteen_flights, nineteen_flights$OP_UNIQUE_CARRIER == "NK")
Spirit_flights_nineteen <- nrow(Spirit_nineteen_data)

# Summer of 2020
Delta_twenty_data <-
  subset(twenty_flights, twenty_flights$OP_UNIQUE_CARRIER == "DL")
Delta_flights_twenty <- nrow(Delta_twenty_data)

United_twenty_data <-
  subset(twenty_flights, twenty_flights$OP_UNIQUE_CARRIER == "UA")
United_flights_twenty <- nrow(United_twenty_data)
```

```
American_twenty_data <-
  subset(twenty_flights, twenty_flights$OP_UNIQUE_CARRIER == "AA")
American_flights_twenty <- nrow(American_twenty_data)

Southwest_twenty_data <-
  subset(twenty_flights, twenty_flights$OP_UNIQUE_CARRIER == "WN")
Southwest_flights_twenty <- nrow(Southwest_twenty_data)

Spirit_twenty_data <-
  subset(twenty_flights, twenty_flights$OP_UNIQUE_CARRIER == "NK")
Spirit_flights_twenty <- nrow(Spirit_twenty_data)

# Data we are considering
flights_considered_nineteen_data <-
  rbind(Delta_nineteen_data, United_nineteen_data, American_nineteen_data,
        Southwest_nineteen_data, Spirit_nineteen_data)

flights_considered_nineteen <-
  c(Delta_flights_nineteen, United_flights_nineteen, American_flights_nineteen,
    Southwest_flights_nineteen, Spirit_flights_nineteen)

flights_considered_twenty_data <-
  rbind(Delta_twenty_data, United_twenty_data, American_twenty_data,
        Southwest_twenty_data, Spirit_nineteen_data)

flights_considered_twenty <-
  c(Delta_flights_twenty, United_flights_twenty, American_flights_twenty,
```

```
  Southwest_flights_twenty, Spirit_flights_twenty)

# Formatting the date
flights_considered_nineteen_data$Date <-
  as.Date(with(flights_considered_nineteen_data,
               paste(2019, MONTH, DAY_OF_MONTH,sep="-")),"%Y-%m-%d")

flights_considered_twenty_data$Date <-
  as.Date(with(flights_considered_twenty_data,
               paste(2020, MONTH, DAY_OF_MONTH, sep = "-")), "%Y-%m-%d")

# airlines we are considering
airlines <- c("Delta", "United", "American", "Southwest", "Spirit")

# Data cleaning for cancellations between summer of 2019 and summer of 2020
nineteen_cancellation_data <-
  subset(flights_considered_nineteen_data,
         flights_considered_nineteen_data$CANCELLED == 1)
nineteen_cancellations <- nrow(nineteen_cancellation_data)

twenty_cancellation_data <-
  subset(flights_considered_twenty_data,
         flights_considered_twenty_data$CANCELLED == 1)
twenty_cancellations <- nrow(twenty_cancellation_data)

Delta_cancelled_19_data <-
  subset(Delta_nineteen_data, Delta_nineteen_data$CANCELLED == 1)
United_cancelled_19_data <-
```

```
  subset(United_nineteen_data, United_nineteen_data$CANCELLED == 1)
American_cancelled_19_data <-
  subset(American_nineteen_data, American_nineteen_data$CANCELLED == 1)
Southwest_cancelled_19_data <-
  subset(Southwest_nineteen_data, Southwest_nineteen_data$CANCELLED == 1)
Spirit_cancelled_19_data <-
  subset(Spirit_nineteen_data, Spirit_nineteen_data$CANCELLED == 1)

Delta_cancelled_20_data <-
  subset(Delta_twenty_data, Delta_twenty_data$CANCELLED == 1)
United_cancelled_20_data <-
  subset(United_twenty_data, United_twenty_data$CANCELLED == 1)
American_cancelled_20_data <-
  subset(American_twenty_data, American_twenty_data$CANCELLED == 1)
Southwest_cancelled_20_data <-
  subset(Southwest_twenty_data, Southwest_twenty_data$CANCELLED == 1)
Spirit_cancelled_20_data <-
  subset(Spirit_twenty_data, Spirit_twenty_data$CANCELLED == 1)

cancelled_19 <- c(nrow(Delta_cancelled_19_data),
                  nrow(United_cancelled_19_data),
                  nrow(American_cancelled_19_data),
                  nrow(Southwest_cancelled_19_data),
                  nrow(Spirit_cancelled_19_data))

cancelled_20 <- c(nrow(Delta_cancelled_20_data),
                  nrow(United_cancelled_20_data),
                  nrow(American_cancelled_20_data),
```

```
                 nrow(Southwest_cancelled_20_data),
                 nrow(Spirit_cancelled_20_data))

all_cancelled <- cbind(cancelled_19, cancelled_20)

Delta_19_cancellation_ratio <- (nrow(Delta_cancelled_19_data) /
                                nrow(Delta_nineteen_data))
Delta_19_cancellation_percent <- (Delta_19_cancellation_ratio * 100)

United_19_cancellation_ratio <- (nrow(United_cancelled_19_data) /
                                 nrow(United_nineteen_data))
United_19_cancellation_percent <- (United_19_cancellation_ratio * 100)

American_19_cancellation_ratio <- (nrow(American_cancelled_19_data) /
                                   nrow(American_nineteen_data))
American_19_cancellation_percent <- (American_19_cancellation_ratio * 100)

Southwest_19_cancellation_ratio <- (nrow(Southwest_cancelled_19_data) /
                                    nrow(Southwest_nineteen_data))
Southwest_19_cancellation_percent <- (Southwest_19_cancellation_ratio * 100)

Spirit_19_cancellation_ratio <- (nrow(Spirit_cancelled_19_data) /
                                 nrow(Spirit_nineteen_data))
Spirit_19_cancellation_percent <- (Spirit_19_cancellation_ratio * 100)


Delta_20_cancellation_ratio <- (nrow(Delta_cancelled_20_data) /
                                nrow(Delta_twenty_data))
```

```
Delta_20_cancellation_percent <- (Delta_20_cancellation_ratio * 100)

United_20_cancellation_ratio <- (nrow(United_cancelled_20_data) /
                                 nrow(United_twenty_data))
United_20_cancellation_percent <- (United_20_cancellation_ratio * 100)

American_20_cancellation_ratio <- (nrow(American_cancelled_20_data) /
                                   nrow(American_twenty_data))
American_20_cancellation_percent <- (American_20_cancellation_ratio * 100)

Southwest_20_cancellation_ratio <- (nrow(Southwest_cancelled_20_data) /
                                    nrow(Southwest_twenty_data))
Southwest_20_cancellation_percent <- (Southwest_20_cancellation_ratio * 100)

Spirit_20_cancellation_ratio <- (nrow(Spirit_cancelled_20_data) /
                                 nrow(Spirit_twenty_data))
Spirit_20_cancellation_percent <- (Spirit_20_cancellation_ratio * 100)


# CODE A - CANCELLED DUE TO CARRIER
cancellation_reason_A_19_data <-
  subset(nineteen_cancellation_data,
         nineteen_cancellation_data$CANCELLATION_CODE == "A" )
cancellation_reason_A_19 <- nrow(cancellation_reason_A_19_data)

# CODE B - CANCELLED DUE TO WEATHER
cancellation_reason_B_19_data <-
  subset(nineteen_cancellation_data,
```

```
        nineteen_cancellation_data$CANCELLATION_CODE == "B" )
cancellation_reason_B_19 <- nrow(cancellation_reason_B_19_data)

# CODE C - CANCELLED DUE TO NATIONAL AIR SYSTEM
cancellation_reason_C_19_data <-
subset(nineteen_cancellation_data,
      nineteen_cancellation_data$CANCELLATION_CODE == "C" )
cancellation_reason_C_19 <- nrow(cancellation_reason_C_19_data)

# CODE D - CANCELLED DUE TO SECUIRTY ISSUES
cancellation_reason_D_19_data <-
  subset(nineteen_cancellation_data,
        nineteen_cancellation_data$CANCELLATION_CODE == "D" )
cancellation_reason_D_19 <- nrow(cancellation_reason_D_19_data)



# CREATING SUBSETS BASED ON REASON FOR CANCELLED (CANCELLATION CODE) 2020
# unique(Delta_cancelled_20_data$CANCELLATION_CODE)

# CODE A - CANCELLED DUE TO CARRIER
cancellation_reason_A_20_data <-
  subset(twenty_cancellation_data,
        twenty_cancellation_data$CANCELLATION_CODE == "A" )
cancellation_reason_A_20 <- nrow(cancellation_reason_A_20_data)

# CODE B - CANCELLED DUE TO WEATHER
cancellation_reason_B_20_data <-
  subset(twenty_cancellation_data,
```

```
        twenty_cancellation_data$CANCELLATION_CODE == "B" )
cancellation_reason_B_20 <- nrow(cancellation_reason_B_20_data)

# CODE C - CANCELLED DUE TO NATIONAL AIR SYSTEM
cancellation_reason_C_20_data <-
  subset(twenty_cancellation_data,
        twenty_cancellation_data$CANCELLATION_CODE == "C" )
cancellation_reason_C_20 <- nrow(cancellation_reason_C_20_data)

# CODE D - CANCELLED DUE TO SECUIRTY ISSUES
cancellation_reason_D_20_data <-
  subset(twenty_cancellation_data,
        twenty_cancellation_data$CANCELLATION_CODE == "D" )
cancellation_reason_D_20 <- nrow(cancellation_reason_D_20_data)



# DATA TO BE CONSIDERED FOR CANCELLATIONS
cancellation_considered_nineteen_data <-
  rbind(cancellation_reason_A_19_data,
        cancellation_reason_B_19_data,
        cancellation_reason_C_19_data,
        cancellation_reason_D_19_data)

cancellation_considered_nineteen_data$Date <-
  as.Date(with(cancellation_considered_nineteen_data,
              paste(2019, MONTH, DAY_OF_MONTH,sep="-")), "%Y-%m-%d")

cancellation_considered_nineteen <-
```

```
  c(cancellation_reason_A_19,
    cancellation_reason_B_19,
    cancellation_reason_C_19,
    cancellation_reason_D_19)

cancellation_considered_twenty_data <-
  rbind(cancellation_reason_A_20_data,
        cancellation_reason_B_20_data,
        cancellation_reason_C_20_data,
        cancellation_reason_D_20_data)
cancellation_considered_twenty_data$Date <-
  as.Date(with(cancellation_considered_twenty_data,
               paste(2020, MONTH, DAY_OF_MONTH,sep="-")), "%Y-%m-%d")

cancellation_considered_twenty <-
  c(cancellation_reason_A_20,
    cancellation_reason_B_20,
    cancellation_reason_C_20,
    cancellation_reason_D_20)

# REASONS FOR CANCELLATION NAMED
cancellation_reasons <-
  c("Carrier Issues",
    "Weather Risk",
    "National Air System Issue",
    "Security Breach")

popular_city_dep_19 <-
```

```
  subset(flights_considered_nineteen_data,
         flights_considered_nineteen_data$ORIGIN_CITY_NAME ==
           "Chicago, IL" |
           flights_considered_nineteen_data$ORIGIN_CITY_NAME ==
           "Dallas/Fort Worth, TX" |
           flights_considered_nineteen_data$ORIGIN_CITY_NAME ==
           "Los Angeles, CA" |
           flights_considered_nineteen_data$ORIGIN_CITY_NAME ==
           "New York, NY")

popular_city_dep_20 <-
  subset(flights_considered_twenty_data,
         flights_considered_twenty_data$ORIGIN_CITY_NAME ==
  "Chicago, IL" |
           flights_considered_twenty_data$ORIGIN_CITY_NAME ==
  "Dallas/Fort Worth, TX" |
           flights_considered_twenty_data$ORIGIN_CITY_NAME ==
  "Los Angeles, CA" |
           flights_considered_twenty_data$ORIGIN_CITY_NAME ==
  "New York, NY")
```

Code with all the plots

```{r}
library(ggplot2)

# ggplot line graph of total flights considered 2019
nineteen_line_gg <- ggplot(flights_considered_nineteen_data) +
```

```
  geom_line(aes(x = Date, y = ..count..), stat = "bin", binwidth = 1) +
  scale_x_date(limits = as.Date(c("2019-06-01", "2019-09-01"))) +
  ylim(0, 12500) +
  ggtitle("Total Flights Considered From Summer 2019")
(nineteen_line_gg)

# ggplot line graph of total flights considered 2020
twenty_line_gg <- ggplot(flights_considered_twenty_data) +
  geom_line(aes(x = Date, y = ..count..), stat = "bin", binwidth = 1) +
  scale_x_date(limits = as.Date(c("2020-06-01", "2020-09-01"))) +
  ylim(0,12500) +
  ggtitle("Total Flights Considered From Summer 2020")
(twenty_line_gg)

# ggplot bar graph of total flights in 2019 for the airlines we are considering
nineteen_bar_gg <- ggplot(flights_considered_nineteen_data) +
  geom_bar(aes(x = factor(OP_UNIQUE_CARRIER,
                         levels = c('AA', 'DL', 'WN', 'NK', 'UA')),
               y = ..count..,
               fill = OP_UNIQUE_CARRIER)) +
  scale_x_discrete(labels = c("AA" = "American",
                              "DL" = "Delta",
                              "WN" = "Southwest",
                              "NK" = "Spirit",
                              "UA" = "United")) +
  ylim(0, 300000) +
  xlab("Airlines") +
  ylab("FLight Count") +
```

```
  ggtitle("Flights in Summer of 2019 Seperated by the Airlines Considered")
(nineteen_bar_gg)


# ggplot bar graph of total flights in 2020 for the airlines we are considering
twenty_bar_gg <- ggplot(flights_considered_twenty_data) +
  geom_bar(aes(x = factor(OP_UNIQUE_CARRIER,
                          levels = c('AA', 'DL', 'WN', 'NK', 'UA')),
               y = ..count..,
               fill = OP_UNIQUE_CARRIER)) +
  scale_x_discrete(labels = c("AA" = "American",
                              "DL" = "Delta",
                              "WN" = "Southwest",
                              "NK" = "Spirit",
                              "UA" = "United")) +
  ylim(0, 300000) +
  xlab("Airlines") +
  ylab("FLight Count") +
  ggtitle("Flights in Summer of 2020 Seperated by the Airlines Considered") +
  labs(color = "Airlines")
twenty_bar_gg


# ggplot boxplot of flight cancellations by airline
cancellation_19_box <- ggplot(data = cancellation_considered_nineteen_data) +
  geom_boxplot(aes(x = factor(OP_UNIQUE_CARRIER,
                              levels = c('AA', 'DL', 'WN', 'NK', 'UA')),
                   y = which(CANCELLED == 1),
                   fill = OP_UNIQUE_CARRIER)) +
  scale_x_discrete(labels = c("AA" = "American",
```

```
                    "DL" = "Delta",
                    "WN" = "Southwest",
                    "NK" = "Spirit",
                    "UA" = "United")) +
  ylim(0, 15000) +
  xlab("Airlines") +
  ylab("Cancellation Count") +
  ggtitle("Cancellation Count by Airline in 2019") +
  labs(color = "Airlines")
cancellation_19_box

cancellation_20_box <- ggplot(data = cancellation_considered_twenty_data) +
  geom_boxplot(aes(x = factor(OP_UNIQUE_CARRIER,
                    levels = c('AA', 'DL', 'WN', 'NK', 'UA')),
               y = which(CANCELLED == 1),
               fill = OP_UNIQUE_CARRIER)) +
  scale_x_discrete(labels = c("AA" = "American",
                    "DL" = "Delta",
                    "WN" = "Southwest",
                    "NK" = "Spirit",
                    "UA" = "United")) +
  ylim(0, 15000) +
  xlab("Airlines") +
  ylab("Cancellation Count") +
  ggtitle("Cancellation Count by Airline in 2020")
cancellation_20_box

# ggplot line plot of departures from popular cities
```

```
departures_19_line <- ggplot(popular_city_dep_19) +
  geom_line(aes(x = Date,
                y = ..count..,
                color = as.factor(ORIGIN_CITY_NAME)),
            stat = "bin",
            binwidth = 1) +
  xlab("Popular Cities") +
  ylab("Departures") +
  ylim(0, 800) +
  ggtitle("Count of Departures from Popular Cities in 2019") +
  labs(color = "Cities")
departures_19_line

departures_20_line <- ggplot(popular_city_dep_20) +
  geom_line(aes(x = Date,
                y = ..count..,
                color = as.factor(ORIGIN_CITY_NAME)),
            stat = "bin",
            binwidth = 1) +
  xlab("Popular Cities") +
  ylab("Departures") +
  ylim(0, 800) +
  ggtitle("Count of Departures from Popular Cities in 2020") +
  labs(color = "Cities")
departures_20_line

# ggplot line graph of the reasons for cancellations
cancellation19_bar_gg <- ggplot(data = cancellation_considered_nineteen_data) +
```

```
  geom_bar(aes(x = factor(CANCELLATION_CODE,
                          levels = c('A', 'B', 'C', 'D')),
               y = ..count..,
              fill = CANCELLATION_CODE)) +
  xlab("Cancellation Reasons") +
  ylab("Count of Cancelled Flights") +
  labs(color = "Cancellation Code") +
  scale_fill_discrete(labels = c("Carrier Issues",
                                 "Weather Risk",
                                 "National Air System Issue",
                                 "Security Breach")) +
  ggtitle("Reasons Flights Were Cancelled in 2019") +
  ylim(0, 7500)
cancellation19_bar_gg

cancellation20_bar_gg <- ggplot(data = cancellation_considered_twenty_data) +
  geom_bar(aes(x = factor(CANCELLATION_CODE,
                          levels = c('A', 'B', 'C', 'D')),
               y = ..count..,
              fill = CANCELLATION_CODE)) +
  xlab("Cancellation Reasons") +
  ylab("Count of Cancelled Flights") +
  labs(color = "Cancellation Code") +
  scale_fill_discrete(labels = c("Carrier Issues",
                                 "Weather Risk",
                                 "National Air System Issue",
                                 "Security Breach")) +
  ggtitle("Reasons Flights Were Canceled in 2020") +
```

```
  ylim(0, 7500)

cancellation20_bar_gg
```

9 - References

- Dataset: https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=FGJ
- https://getawaytips.azcentral.com/busiest-time-air-travel-3409.html
- https://www.tripsavvy.com/busiest-airports-in-the-usa-3301020
- https://www.airlines.org/dataset/u-s-bankruptcies-and-services-cessations/
- https://www.icao.int/Newsroom/Pages/2020-passenger-totals-drop-60-percent-as-COVID19-assault-on-international-mobility-continues.aspx