Alaska Airlines Flight Path Data Analysis in R

Farhan Billah, Jonathan Bowers, & Justin Chow

ECON 43 - Data Wrangling & Visualization

Professor Sunghun Chung

Abstract

In this project, we analyze airline and airport data over 18 years in order to identify and understand flight path demand trends, before the pandemic and during the pandemic. After conducting thorough analysis using several R functions, we have recommended strategic business decisions for Alaska Airlines based on airport growth during and before the pandemic.

Motivation

The United States Department of Transportation's Bureau of Labor Statistics has expansive public data on airlines, airports, flight paths and passenger numbers. Conducting thorough analysis on this data can provide airlines with insights which can be used to make strategic business decisions. Our objective was to take flight path data on 9 major airports and use several R functions to clean, analyze, and visualize the data. We focused on Alaska Airlines, the 16th largest airline by market capitalization, as they are a healthy company with potential for growth. We will form strategic recommendations to optimize Alaska Airlines' flight paths in order to maximize profit and increase market share of flights at airports showing signs of growth.

Data

The dataset used for this project was taken from the United States Department of Transportation's Bureau of Transportation Statistics. It contains data on 14 U.S. and foreign airline companies as well as several airports within the United States. The dataset contains information on several variables from October 2002 to November 2020. The variables include passengers, flights, net income, operating revenue and more. There is data on each of these variables separated by month as well as domestic or international flights. We randomly selected 10 major airports that Alaska Airlines flies to within the United States and downloaded the passenger and flight data for each. The airports included were in Portland, Salt Lake City, Honolulu, Denver, Houston, Orlando, Minneapolis, Detroit, Baltimore, and Philadelphia. Alaska Airlines is headquartered in Seattle, Washington so there is considerably more travel to and from Portland than any other airport due to its proximity to Seattle. Because of this, we elected to omit Portland from the majority of our analysis as it significantly skewed the data. Alaska already has a prominent presence at PDX, so we decided to focus on flight demand for the other airports.

Analysis

By visualizing our data, we were able to draw several conclusions about Alaska Airlines history since 2002 as well as their response to the COVID-19 Pandemic in 2020. When analyzing the flight counts for all airlines in the below line graph (Figure 1), we identified a few trends that would be beneficial in formulating our strategy recommendations. We can see clearly that Denver not only has the highest flight count of these airports, but it is also the only airport to have a significant increase in flights in the last five years. Between 2015 and 2019, Denver International saw a 28% increase in total flights, yet Alaska has only increased flights there by 9% in the same time frame as shown in Figure 2. Alaska has also proportionally increased flight paths to Orlando and Philadelphia in accordance with the increase in total airline flights to those airports, showing us Alaska has experience with quickly negotiating deals with airports as soon as they notice increased demand. Figure 2 also demonstrates Alaska's growing commitment to Hawaii over the past 13 years. Honolulu has become one of Alaska's primary destinations, as evidenced by the steep increase in flights from there between 2007 and 2013. Another interesting finding was that Alaska significantly decreased the amount of flights going in and out of Salt Lake City in the past few years even though there were incremental increases in the number of flights from all airlines to that airport. This means that after substantially increasing their presence in SLC from 2010 to 2016, Alaska chose to change course and decrease their presence there despite the relatively steady, and sometimes rising, overall demand at that airport shown in Figures 1 and 3. Figures 3 and 4 show passengers rather than flights for all major airlines and just Alaska Airlines respectively. They display the same patterns as the flight plots.

Figure 1:

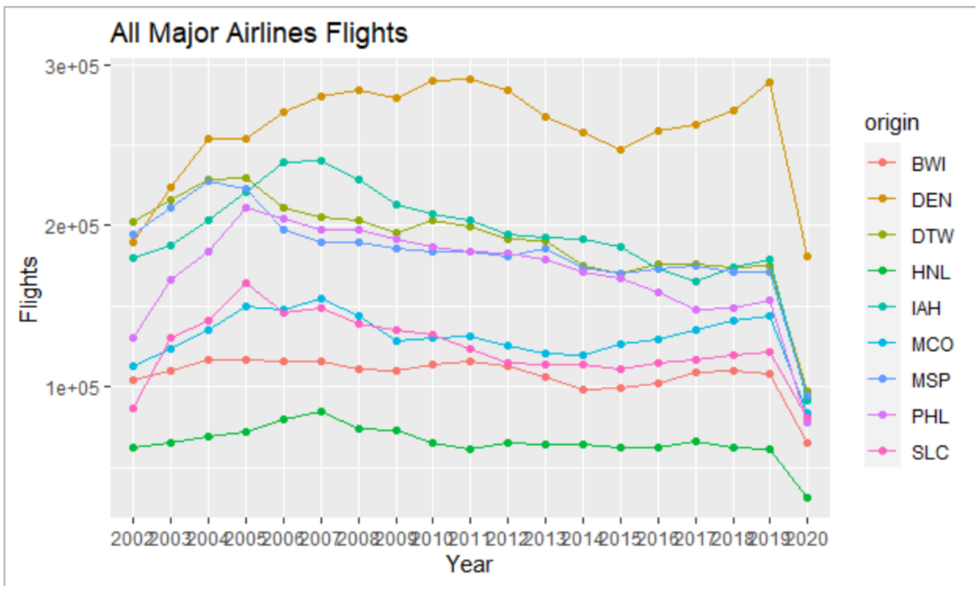


Figure 2:

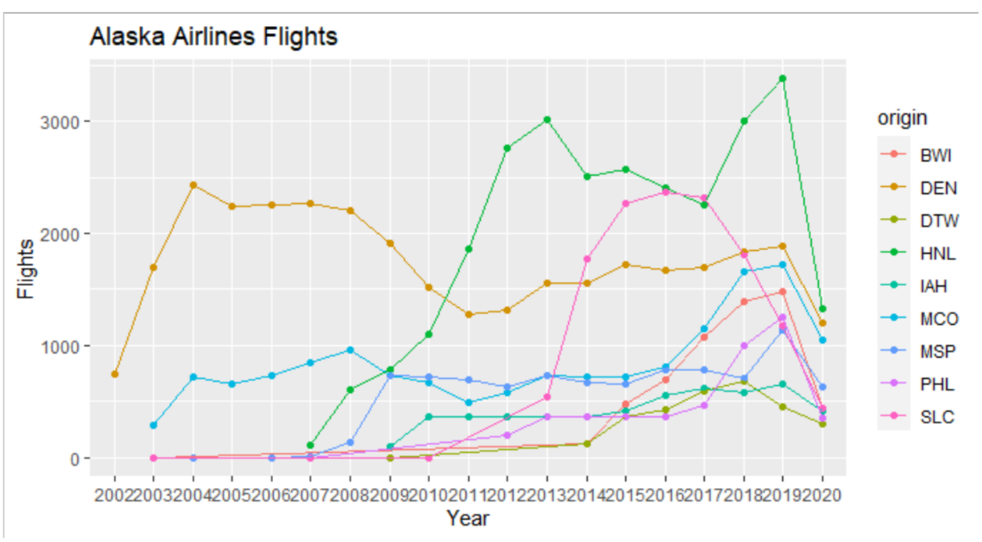


Figure 3:

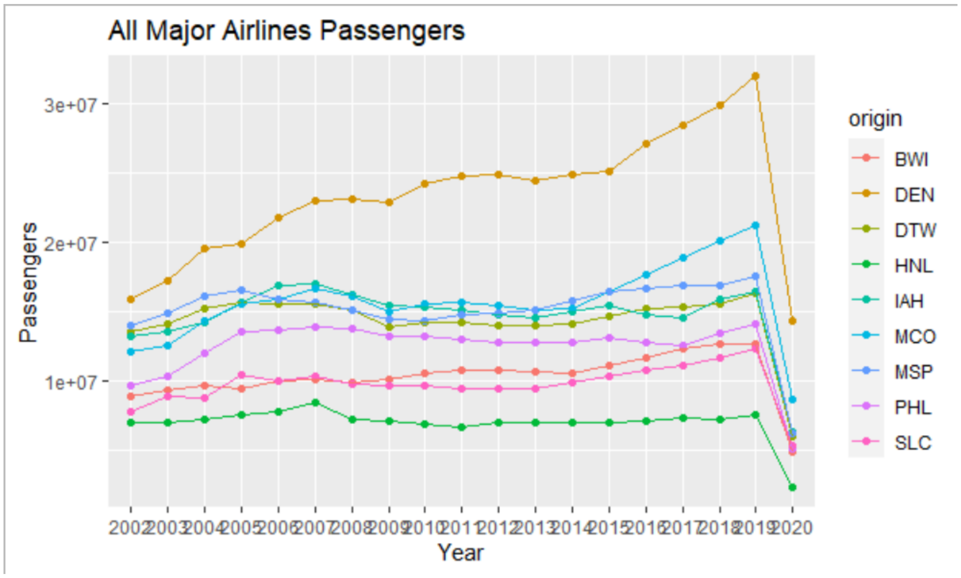
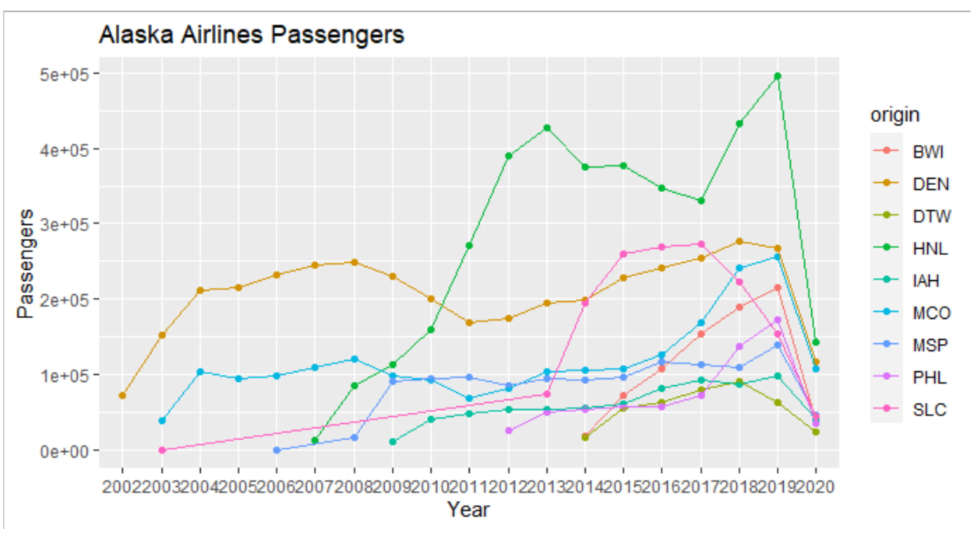


Figure 4:



The COVID-19 Pandemic effectively shut down travel for nearly nine months of 2020 as safety regulations were put in place that barred interstate travel, and people cancelled plans out of fear of contracting the virus. The travel industry responded by downsizing significantly, laying off employees and reducing their output in order to cut costs. The last year of Figures 1-4 show the huge dip in the number of flights and passengers in 2020. Figure 5 shows that in more detail for Alaska Airlines. Passengers from Baltimore and Philadelphia declined by more than 80% and flights were reduced by approximately 70%. Flights from Alaska's top destination in our dataset, Honolulu, were reduced by 61% while flights from the overall busiest airport in our dataset, Denver, were reduced by 36%. However, Figure 5 shows how Alaska's response of cutting flights was uneven with the percentage decrease in passengers. For instance, passengers from Detroit declined by 63% in 2020 but flights were only reduced by about half that, 32%. Figure 6 visualizes this relationship in a different way. The blue line represents the linear trend of the passenger to flight ratio between 2002 and 2019, or before the pandemic. The dots represent the passenger to flight ratio for each airport during 2020. All the points are above the trend line, showing that Alaska did not reduce their flights proportionally with the decrease in demand. The graph implies that in 2020 planes had far more empty seats than in past years. This is to be expected with the reduced demand and could be due to any number of logistical or scheduling reasons. It could also be due to the need for social distancing on planes. Many airlines including Alaska blocked off their middle seats in order to allow for more space between passengers. Flying only partially full airplanes was quite costly for Alaska, and in January 2021, they discontinued their empty middle seat policy claiming that wearing a mask was sufficient protection even with the middle seat occupied.

Figure 5:

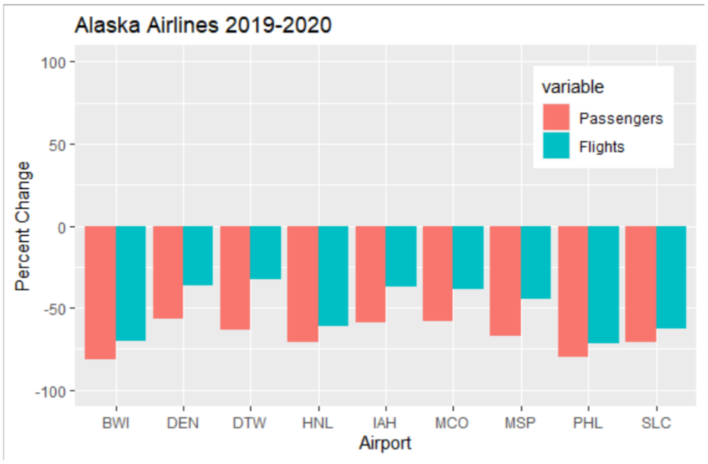
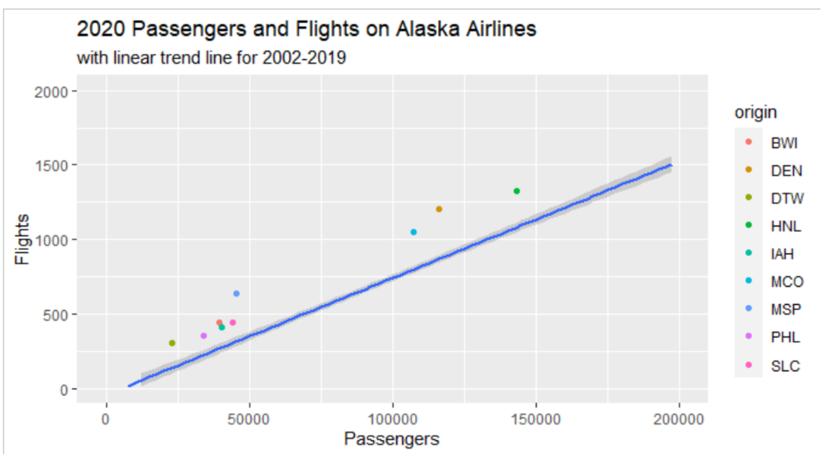


Figure 6:



Results

Measures of Variability:

25th Quartile 75th Quartile	`## 25% 75%` `## 7429.0 31536.5`
Interquartile Range (IQR)	`## [1] 24107.5`
Variance	`## [1] 29975863921`
Summary Statistics	`## Min. 1st Qu. Median Mean 3rd Qu. Max.` `## 48 7429 12632 53949 31537 2529551`

Using only the range of passengers spanning the middle half of the major dataset, we can see that the 25th quartile has 7,429 passengers and the 75th quartile has 31,537 flights. Thus, using the quartile results, we can calculate the interquartile range (IQR) to be 24,107 passengers. To further measure how dispersed the data values are around the mean, we calculate the variance of passengers traveling to major airports to be overwhelmingly positive, which indicates that the data points are very spread out from the mean. To determine whether passengers are affected by skewness, we can use the summary statistics to calculate the median, 18,013 passengers, and the mean, 324,600 passengers. Since the mean of the data set is much larger than the median, we can conclude that the data values are right- or positively-skewed. As a result, it is safe to assume that Alaska Airlines did not reduce the number of flights proportionately with decrease in passengers, and their mean number of passengers is heavily skewed by other significant variables.

Descriptive Statistics:

Descriptive Statistics for Alaska Airlines passengers to Major Airports

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
domestic	422	13,697.8	10,030.9	48	5,942.5	11,031	20,050.8	47,423
international	3	285.0	53.7	227.0	261.0	295.0	314.0	333.0
total	396	14,298.7	10,033.3	48.0	6,959.0	11,792.5	20,652.0	47,423.0

Using a descriptive statistics table, we can subset unnecessary missing values as well as omit Portland airport as it significantly skews the data. To further reduce skewness in our analysis, we analyse passengers that were recorded on flights after 2017. This subset allows our group to analyze the effect of COVID-19 pandemic on mean and median numbers of passengers. Most notably the mean number of passengers drastically decreased by about 40,000 and median number of flights decreased by about 1,000. Another significant drop can be seen in the 25th and 75th percentile, where the number of passengers dropped to about 6,000 and 20,000 passengers respectively. These results paired with the fact that we chose to omit Portland's major airport shows how reliant Alaska Airlines is on Portland airport passengers. Overall, our group concluded that demand for domestic air travel using Alaska Airlines has greatly decreased.

Regression Analysis:

```
Regression
==============================================
                      Dependent variable:
                   ---------------------------
                         lnpassengers
----------------------------------------------
year2018                    0.140*
                           (0.084)

year2019                    0.162*
                           (0.084)

year2020                  -1.260***
                           (0.086)

originDEN                  0.590***
                           (0.128)

originDTW                 -0.791***
                           (0.127)

originHNL                  0.912***
                           (0.128)

originIAH                 -0.471***
                           (0.128)

originMCO                  0.375***
                           (0.128)

originMSP                  -0.234*
                           (0.128)

originPHL                 -0.346***
                           (0.128)

originSLC                   -0.062
                           (0.129)

Constant                   9.400***
                           (0.104)

----------------------------------------------
Observations                  422
R2                           0.617
Adjusted R2                  0.607
Residual Std. Error          0.619
F Statistic                60.037***
==============================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

In our regression analysis, we used the year and origin variable on log passengers. We can see that year2018 and year2019 is positively associated with log passengers at a 1% significance level, whereas year2020 is negatively associated with log passengers at a 1% significance level. Using the coefficient values, if the year2018 increases by one, passengers increase by 16.7%. However, if the year2020 increases by one, passengers are predicted to decrease by more than 102.6%. Overall, the variable year serves as a strong factor that helps explain the number of passengers. Passengers traveling from Alaska's top destination in our dataset, Honolulu, increased by 91.2% while passengers from the overall busiest airport in our dataset, Denver, increased by more than 59%, and Orlando, was increased by 37.5%. These major airports' positive association with passengers shows that US residents are moving out of major metropolitans to developing economies. Despite the threat of COVID-19 pandemic, these major airports are slightly less but still positively associated with passengers. However, passengers traveling from Alaska's other top destination in our dataset, Houston, was decreased by more than 47% while passengers from other busy airports, Minneapolis, was decreased by more than 23%, Philadelphia, was decreased by more than 34%, and Salt Lake City, was decreased by more than 6%. The lower number of passengers to these major airports means that demand is predicted to reduce to these destinations. Overall, origin is a strong factor that helps explain the number of passengers. Our group determined that using year and origin on passengers, fits well with the linear model from the adjusted-R2 being greater than 80%. The year2020 has a clear negative associated with passengers, but major airports can either be positively or negatively associated with passengers.

Analysis of Variance (ANOVA):

```
             Df Sum Sq Mean Sq F value   Pr(>F)    
year          3 142.48   47.49 146.933  < 2e-16 ***
origin        8 110.93   13.87  42.900  < 2e-16 ***
year:origin  24  32.56    1.36   4.197 6.74e-10 ***
Residuals   386 124.76    0.32                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using the Analysis of Variance (ANOVA), our group noticed that there is a meaningful year difference in the number of passengers. An F-statistic value of 146.933 and p-value less than 0.05 means our sample data provides sufficient evidence to conclude that the regression model fits the data better than a model with no independent variables. Similarly, there is a meaningful origin difference in the number of passengers. An F-statistic value of 42.9 and p-value less than 0.05 means our sample data provides sufficient evidence to conclude that the regression model fits the data better than a model with no independent variables. Overall, the variables year and origin on passengers are considered a good fit. Finally, there is a meaningful difference in interaction between year and origin. An F-statistic value of 4.197 and p-value less than 0.05 means our sample data provides sufficient evidence to conclude that the regression model fits the data better than a model with no independent variables. Compared to year and origin individual impact on passengers, the interaction of year and origin on passengers still serves as a good fit. However, the two interactive categorical variables having all factors being considered at the same time has a higher chance of variability, which results in a weaker fit.

Conclusion

After conducting our analysis and visualizing all of our data, we decided to focus on creating three actionable recommendations we could give to Alaska Airlines based on flight path demand that would optimize their flight paths to maximize profits and increase market share.

The increase in flights to Denver is likely due to the fact that Denver's rate of growth has skyrocketed, as several companies have established themselves in the city and its suburbs, bringing tens of thousands of new residents there every year. We recommend that Alaska increase flights from surrounding states to Denver and vice versa as our results show a considerable increase in passengers in the last five years.

When analyzing Houston passenger trends, we decided to recommend that Alaska increase flight paths to and from Houston as it saw the second smallest percentage decline post-COVID, meaning more people are willing to travel here even amid a pandemic. Houston is a growing city in a state with more lax covid restrictions, so there will be benefits to increasing flights both during and after the pandemic.

Finally, Alaska elected to decrease flights in and out of Salt Lake City, but SLC has seen a healthy incremental increase each year leading up to COVID. Salt Lake City is another western city that has seen astronomical increases in housing prices and corporate activity in the past few years. During COVID, there has been a mass migration of Californians moving to cheaper Western states, meaning there will be more traffic in and out of SLC. Considering its proximity to Alaska's hub Seattle, their regional presence will give them an advantage in taking a large portion of this market share. It is our strategic recommendation that Alaska should negotiate a deal with the airport to secure its spot in this growing hub.

Appendix

Code from R:

```
# Clear the working space
rm(list = ls())

# Set working directory
setwd("C:/Users/j-bow/Documents/ECON 42/data")

library(readxl)
library(ggplot2)
library(reshape)
library(dplyr)

den <- read_xls("DEN.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
den$origin <- "DEN"
bwi <- read_xls("BWI.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
bwi$origin <- "BWI"
dtw <- read_xls("DTW.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
dtw$origin <- "DTW"
slc <- read_xls("SLC.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
slc$origin <- "SLC"
pdx <- read_xls("PDX.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
pdx$origin <- "PDX"
phl <- read_xls("PHL.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
phl$origin <- "PHL"
mco <- read_xls("MCO.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
mco$origin <- "MCO"
msp <- read_xls("MSP.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
msp$origin <- "MSP"
iah <- read_xls("IAH.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
```

```
iah$origin <- "IAH"
hnl <- read_xls("HNL.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
hnl$origin <- "HNL"
flights <- read_xls("Flights_3_1_2021 3_57_01 AM.xls", col_names = c("Year", "Month",
"Domestic", "International", "Total"), skip = 2)
passengers <- read_xls("Passengers_3_1_2021 3_56_53 AM.xls", col_names = c("Year",
"Month", "Domestic", "International", "Total"), skip = 2)

denf <- read_xls("DENflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
denf$origin <- "DEN"
bwif <- read_xls("BWIflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
bwif$origin <- "BWI"
dtwf <- read_xls("DTWflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
dtwf$origin <- "DTW"
slcf <- read_xls("SLCflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
slcf$origin <- "SLC"
pdxf <- read_xls("PDXflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
pdxf$origin <- "PDX"
phlf <- read_xls("PHLflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
phlf$origin <- "PHL"
mcof <- read_xls("MCOflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
mcof$origin <- "MCO"
mspf <- read_xls("MSPflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
mspf$origin <- "MSP"
iahf <- read_xls("IAHflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
iahf$origin <- "IAH"
hnlf <- read_xls("HNLflights.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
hnlf$origin <- "HNL"
```

```
majorp <- rbind(bwi,den,dtw,hnl,iah,mco,msp,pdx,phl,slc)
majorf <- rbind(bwif,denf,dtwf,hnlf,iahf,mcof,mspf,pdxf,phlf,slcf)
total <- merge(majorp, majorf, by=c("Year", "Month", "origin"))

# Data Cleaning
# covert everything to lower case
names(major) <- tolower(names(major))
# drop missing values in major airport dataset
major <- major %>% filter(!is.na(domestic))
# subset major airport dataset
subset1 <- subset(major, year >= 2017 & origin != "PDX" & month != "TOTAL")

# Measure of Variability
quantile(major$domestic, probs = c(0.25, 0.75), na.rm = TRUE)
IQR(major$domestic, na.rm = TRUE)
var(major$domestic)
skew(major$domestic)
kurtosi(major$domestic)
summary(major$domestic)

# Descriptive Statistics
stargazer(subset(as.data.frame(subset1), year >= 2017 & origin != "PDX" & month !=
"TOTAL"), type="text", median = TRUE, digits = 1, title="Descriptive Statistics for Alaska
Airlines passengers to Major Airports")

# Create log flights for major dataset
subset1$lnpassengers <- log(subset1$domestic)
# Regression Analysis
myModal <- lm(lnpassengers ~ year + origin, data=subset1)
stargazer(myModal,
        title="Regression", type="text",
        df=FALSE, digits=3)

# Analysis of Variance (ANOVA)
myANOVA <- aov(lnpassengers ~ year*origin, data=subset1)
summary(myANOVA)

# Plot
ggplot(data=subset(majorp, Month =="TOTAL" & origin != "PDX"), aes(x=Year, y=Domestic,
group=origin, color=origin))+
```

```
  geom_point()+
  geom_line()+
  labs(title="Alaska Airlines Passengers",y="Passengers")

ggplot(data=subset(majorf, Month =="TOTAL" & origin != "PDX"), aes(x=Year,
y=Domestic_Flights, group=origin, color=origin))+
  geom_point()+
  geom_line()+
  labs(title="Alaska Airlines Flights",y="Flights")

subset1 <- subset(total, Month=="TOTAL" & origin != "PDX")
subset2 <- subset(total, Month=="TOTAL" & Year > 2018)
subset3 <- total %>%  filter(Month=="TOTAL" & Year ==2020 & origin != "PDX") %>%
  select(origin, Domestic, Domestic_Flights)
subset3$perc_p <-
-100*(1-(subset1$Domestic[subset1$Year==2020]/subset1$Domestic[subset1$Year==2019]))
subset3$perc_f <-
-100*(1-(subset1$Domestic_Flights[subset1$Year==2020]/subset1$Domestic_Flights[subset1$
Year==2019]))
subset3 <- melt(subset3, id=c("origin", "Domestic", "Domestic_Flights"))
subset4 <- total %>%  filter(Month=="TOTAL" & Year ==2020 & origin != "PDX") %>%
  select(origin, Domestic, Domestic_Flights)
subset4$perc_p <-
100*(subset1$Domestic[subset1$Year==2019]/subset1$Domestic[subset1$Year==2014])
subset4$perc_f <-
100*(subset1$Domestic_Flights[subset1$Year==2019]/subset1$Domestic_Flights[subset1$Year
==2014])
subset4 <- melt(subset4, id=c("origin", "Domestic", "Domestic_Flights"))


ggplot(data=subset3, aes(x=origin, y=value, fill=variable)) +
  geom_bar(stat="identity", position = "dodge")+
  ylim(-100,100)+
  labs(title="Alaska Airlines 2019-2020",x="Airport",y="Percent Change")+
  scale_fill_discrete(labels = c("Passengers", "Flights"))+
  theme(legend.position = c(0.85, 0.8))

ggplot(data=subset4, aes(x=origin, y=value, fill=variable)) +
  geom_bar(stat="identity", position = "dodge")+
  ylim(-100,1500)+
```

```
  labs(title="Alaska Airlines 2014-2019",x="Airport",y="Percent Change")+
  scale_fill_discrete(labels = c("Passengers", "Flights"))+
  theme(legend.position = c(0.85, 0.8))

ggplot(data=subset(majorp, Month =="TOTAL" & origin != "PDX" & Year > 2018), aes(x=Year,
y=Domestic, group=origin, color=origin))+
  geom_point()+
  geom_line()+
  labs(title="Decline in Alaska Airlines Passengers 2019-20",y="Passengers")

ggplot(data=subset(majorf, Month =="TOTAL" & origin != "PDX" & Year > 2018), aes(x=Year,
y=Domestic_Flights, group=origin, color=origin))+
  geom_point()+
  geom_line()+
  labs(title="Decline in Alaska Airlines Flights 2019-20",y="Flights")

ggplot(data=NULL)+
  geom_point(data=subset(total, Month=="TOTAL" & origin != "PDX" & Year== 2018),
aes(x=Domestic, y=Domestic_Flights, color=origin))+
  geom_smooth(data=subset(total, Month=="TOTAL" & origin !="PDX" & Year !=2020),
aes(x=Domestic, y=Domestic_Flights), method=lm)+
  xlim(0,200000)+
  ylim(0,2000)+
  labs(title="2020 Passengers and Flights on Alaska Airlines",subtitle="with linear trend line for
2002-2019",x="Passengers",y="Flights")


den_allp <- read_xls("den_allp.xls", col_names = c("Year", "Month", "Domestic",
"International", "Total"), skip = 2)
den_allp$origin <- "DEN"
bwi_allp <- read_xls("bwi_allp.xls", col_names = c("Year", "Month", "Domestic",
"International", "Total"), skip = 2)
bwi_allp$origin <- "BWI"
dtw_allp <- read_xls("dtw_allp.xls", col_names = c("Year", "Month", "Domestic",
"International", "Total"), skip = 2)
dtw_allp$origin <- "DTW"
slc_allp <- read_xls("slc_allp.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
slc_allp$origin <- "SLC"
```

```
pdx_allp <- read_xls("pdx_allp.xls", col_names = c("Year", "Month", "Domestic",
"International", "Total"), skip = 2)
pdx_allp$origin <- "PDX"
phl_allp <- read_xls("phl_allp.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
phl_allp$origin <- "PHL"
mco_allp <- read_xls("mco_allp.xls", col_names = c("Year", "Month", "Domestic",
"International", "Total"), skip = 2)
mco_allp$origin <- "MCO"
msp_allp <- read_xls("msp_allp.xls", col_names = c("Year", "Month", "Domestic",
"International", "Total"), skip = 2)
msp_allp$origin <- "MSP"
iah_allp <- read_xls("iah_allp.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
iah_allp$origin <- "IAH"
hnl_allp <- read_xls("hnl_allp.xls", col_names = c("Year", "Month", "Domestic", "International",
"Total"), skip = 2)
hnl_allp$origin <- "HNL"

den_all <- read_xls("den_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
den_all$origin <- "DEN"
bwi_all <- read_xls("bwi_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
bwi_all$origin <- "BWI"
dtw_all <- read_xls("dtw_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
dtw_all$origin <- "DTW"
slc_all <- read_xls("slc_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
slc_all$origin <- "SLC"
pdx_all <- read_xls("pdx_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
pdx_all$origin <- "PDX"
phl_all <- read_xls("phl_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
phl_all$origin <- "PHL"
mco_all <- read_xls("mco_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
mco_all$origin <- "MCO"
```

```
msp_all <- read_xls("msp_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
msp_all$origin <- "MSP"
iah_all <- read_xls("iah_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
iah_all$origin <- "IAH"
hnl_all <- read_xls("hnl_all.xls", col_names = c("Year", "Month", "Domestic_Flights",
"International_Flights", "Total_Flights"), skip = 2)
hnl_all$origin <- "HNL"

majorp2 <-
rbind(bwi_allp,den_allp,dtw_allp,hnl_allp,iah_allp,mco_allp,msp_allp,pdx_allp,phl_allp,slc_allp
)
majorf2 <- rbind(bwi_all,den_all,dtw_all,hnl_all,iah_all,mco_all,msp_all,pdx_all,phl_all,slc_all)
total2 <- merge(majorp2, majorf2, by=c("Year", "Month", "origin"))

ggplot(data=subset(majorp2, Month =="TOTAL" & origin != "PDX"), aes(x=Year, y=Domestic,
group=origin, color=origin))+
  geom_point()+
  geom_line()+
  labs(title="All Major Airlines Passengers",y="Passengers")

ggplot(data=subset(majorf2, Month =="TOTAL" & origin != "PDX"), aes(x=Year,
y=Domestic_Flights, group=origin, color=origin))+
  geom_point()+
  geom_line()+
  labs(title="All Major Airlines Flights",y="Flights")
```