TABLE OF CONTENTS

01

Project & Data Description

Data Description

Data Source	Actual flight information reported by certified United States airlines to the Bureau of Transportation Statistics	
About the Datasets	**Date Range**	January 2015 - December 2019
	Structure	>1 workbooks (Flight Datasets, Airports.csv, Reporting_Airline.csv) Flight Datasets: one workbook for each year each month (60 workbooks in total)

Data Description

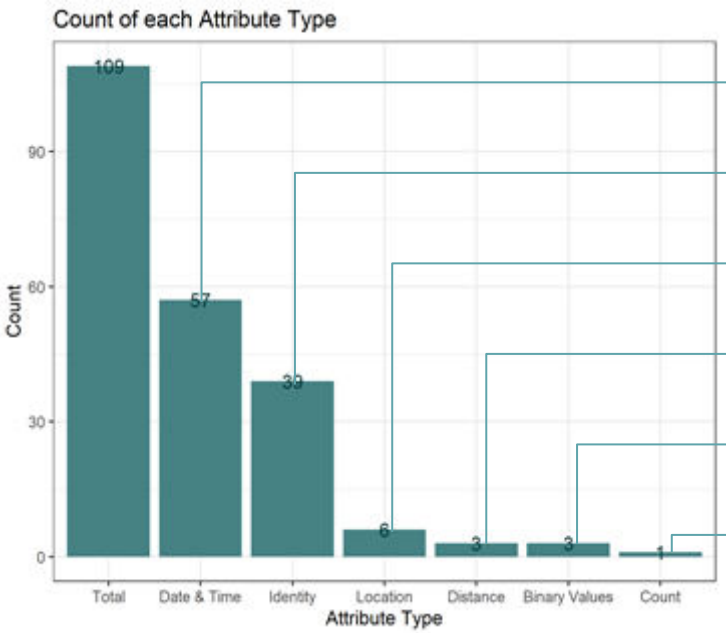
Flight_Data_2015_1.csv
Flight_Data_2015_2.csv
Flight_Data_2015_3.csv
⋮
Flight_Data_2019_12.csv

Airports.csv

Reporting_Airline.csv

109 variables (5,388,873 rows of record) :

- 59 categorical variables
- 50 ordered variables



FlightDate, DepTime, ArrTime, ...

Reporting_Airline, OriginAirportID, CancellationCode, ...

Origin, OriginCityName, OriginState, ...

Distance, DistanceGroup, DivDistance

Cancelled, Diverted, ArrDel15

DivAirportLandings

Data Description

Flight_Data_2015_1.csv
Flight_Data_2015_2.csv
Flight_Data_2015_3.csv
⋮
Flight_Data_2019_12.csv

Airports.csv

Reporting_Airline.csv

7 variables (322 rows of record) :

- 5 categorical variables
- 2 ordered variables

Variable Name	Definition
IATA_CODE	Code assigned by IATA and commonly used to identify a carrier. As the same code may have been assigned to different carriers over time, the code is not always unique. For analysis, use the Unique Carrier Code.
AIRPORT	Airport name
CITY	Airport located city name
STATE	Airport located state name in short form
COUNTRY	Airport located country name in short form
LATITUDE	Latitude of the airport
LONGITUDE	Longitude of the airport

Data Description

Flight_Data_2015_1.csv
Flight_Data_2015_2.csv
Flight_Data_2015_3.csv
⋮
Flight_Data_2019_12.csv

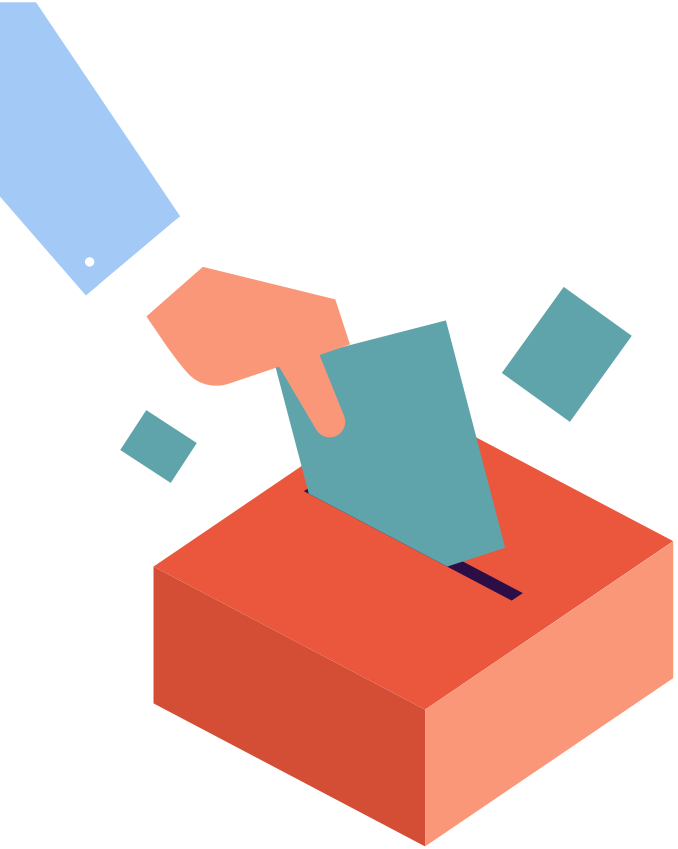
Airports.csv

Reporting_Airline.csv

2 variables (1,665 rows of record) :

- 2 categorical variables
- 0 ordered variables

Variable Name	Definition
Code	Airline Code
Description	Airline Name



02

Expected Findings

Project Description

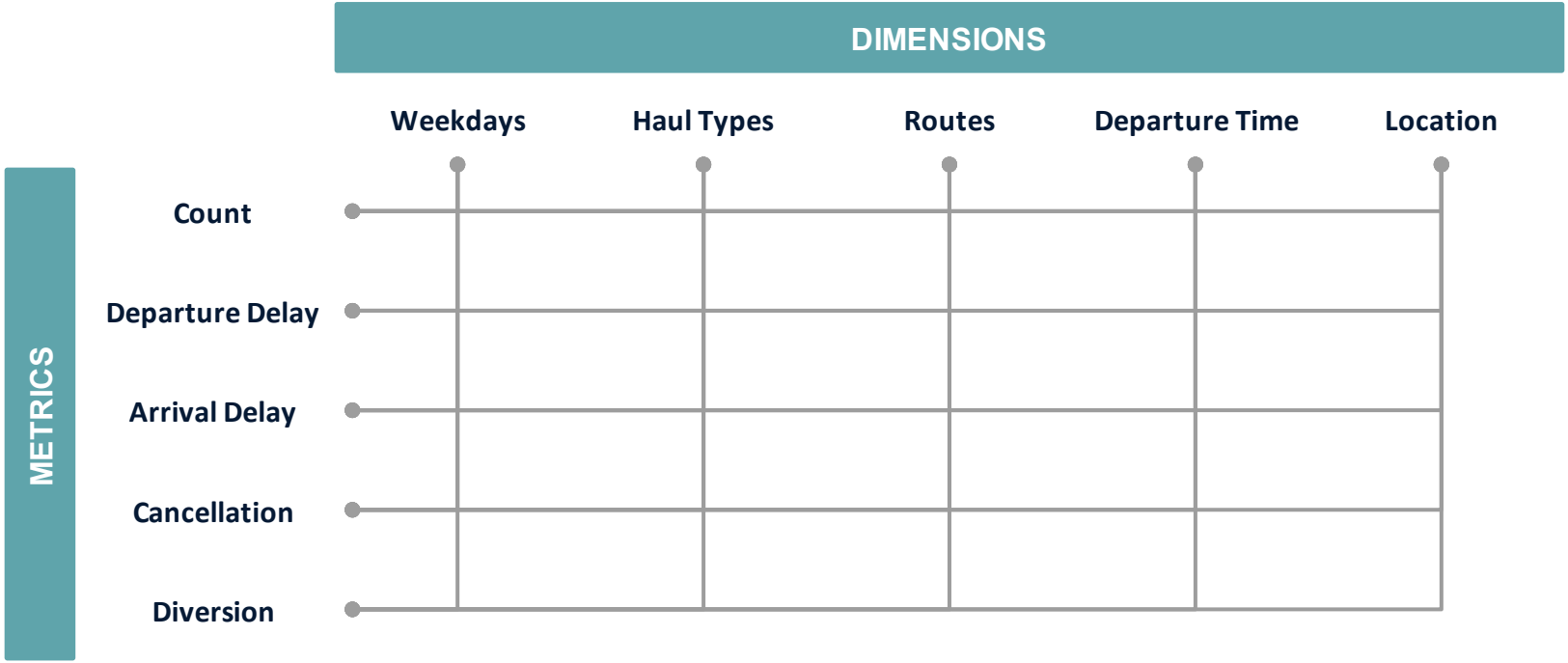
Project Goal Analyzing each airline and identifying top players' performances

Motivation

As a customer Which airline should be the most trustworthy?

Which airline is most suitable for customers like me?

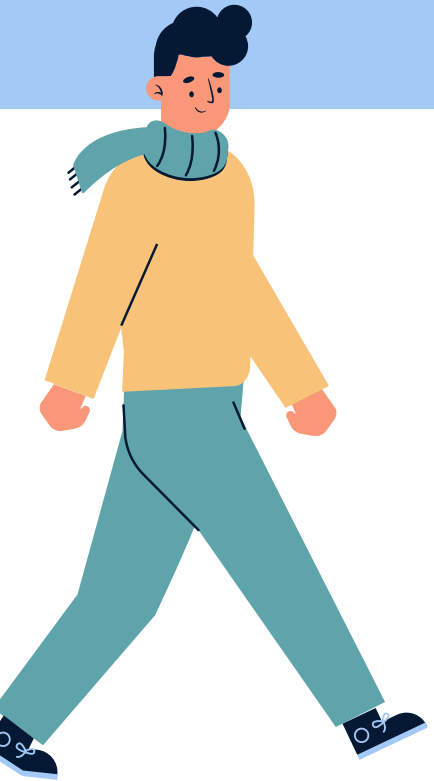
Performance Metrics



Expected Findings

- Popular routes are dominated by top players
- Most long-haul flights are provided by top players
- Top players have less delay, cancellation and diverted flights
- Some airlines specialise in certain locations and dominate that market(s)
- Uncovered factors that may affect the performance of an airline

03



Steps of Analysis

Steps of Analysis

Step1: Data Merging

Data Merging

Flight_Data_2015_1.csv
Flight_Data_2015_2.csv
Flight_Data_2015_3.csv
⋮
Flight_Data_2019_12.csv

Airports.csv

Reporting_Airline.csv

Union Join
[60 Flight Datasets]

Left Join [by Origin Airport]

Left Join [by Destination Airport]

Left Join [by Airline Code]

Step2: Data Cleaning

Data Cleaning

Flight_Data_2015_1.csv
Flight_Data_2015_2.csv
Flight_Data_2015_3.csv
⋮
Flight_Data_2019_12.csv

Airports.csv

Reporting_Airline.csv

Identify which dataset requires data cleaning

1. Are there unnecessary variables?
2. Are there duplications in records?
3. Are there missing values?

Data Cleaning

	Are there unnecessary variables?	Are there duplications in records?	Are there missing values?
Flight Datasets	✔	✘	✔
Airports Dataset	✘	✘	✘
Reporting Airlines Dataset	✘	✘	✘

Data Cleaning

	Are there unnecessary variables?	Are there duplications in records?	Are there missing values?
Flight Datasets	✔	✘	✔
Airports Dataset	✘	✘	✘
Reporting Airlines Dataset	✘	✘	✘

Only Flight Dataset Requires Data Cleaning

Data Cleaning (Variable Selection - 1-13/25)

Variable Name	Definition
Year	Year
Month	Month
DayOfWeek	Day of Week
IATA_CODE_Reporting_Airline	Code assigned by IATA and commonly used to identify a carrier. As the same code may have been assigned to different carriers over time, the code is not always unique. For analysis, use the Unique Carrier Code.
Origin	Origin Airport
OriginCityName	Origin Airport, City Name
OriginStateName	Origin Airport, State Name

Variable Name	Definition
Dest	Destination Airport
DestCityName	Destination Airport, City Name
DestStateName	Destination Airport, State Name
DepTime	Actual Departure Time (local time: hhmm)
DepDelay	Difference in minutes between scheduled and actual departure time. Early departures show negative numbers.
DepDel15	Departure Delay Indicator, 15 Minutes or More (1=Yes)

Data Cleaning (Variable Selection - 15-25/25)

Variable Name	Definition
ArrTime	Actual Arrival Time (local time: hhmm)
ArrDelay	Difference in minutes between scheduled and actual arrival time. Early arrivals show negative numbers.
ArrDel15	Arrival Delay Indicator, 15 Minutes or More (1=Yes)
Cancelled	Cancelled Flight Indicator (1=Yes)
CancellationCode	Specifies The Reason For Cancellation
Diverted	Diverted Flight Indicator (1=Yes)

Variable Name	Definition
Distance	Distance between airports (miles)
CarrierDelay	Carrier Delay, in Minutes
WeatherDelay	Weather Delay, in Minutes
NASDelay	National Air System Delay, in Minutes
SecurityDelay	Security Delay, in Minutes
LateAircraftDelay	Late Aircraft Delay, in Minutes

Data Cleaning (Missing Value Handling)

1. Replace all missing values in CancellationCode by “NA”
2. Replace all missing values in CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay by 0
3. Remove other records with one or above missing values

Step3: Data Visualization

Data Visualization

Tableau:

- Area Charts
- Bar Charts
- Destination Map
- Line Charts
- Packed Bubbles
- Pie Charts
- Scatter Plots
- Stacked Bars
- Treemap

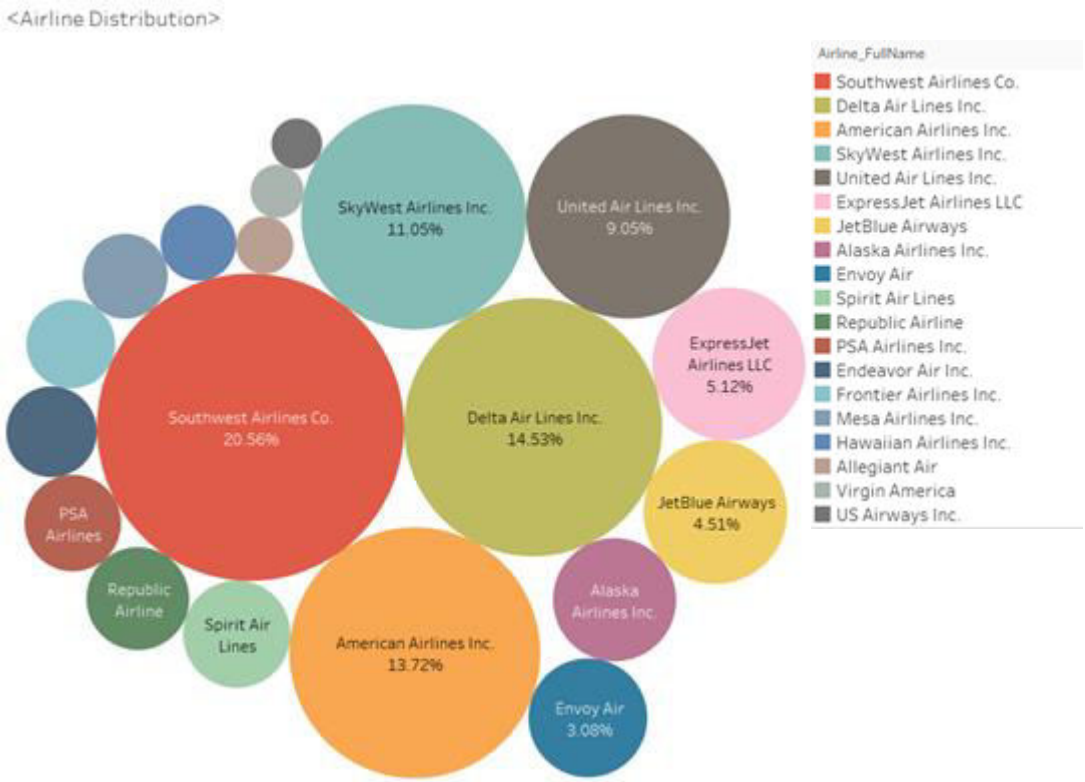
R (Package used):

- corrplot
- plotly
- tidyverse(including ggplot2)
- usmap

04

Main Findings

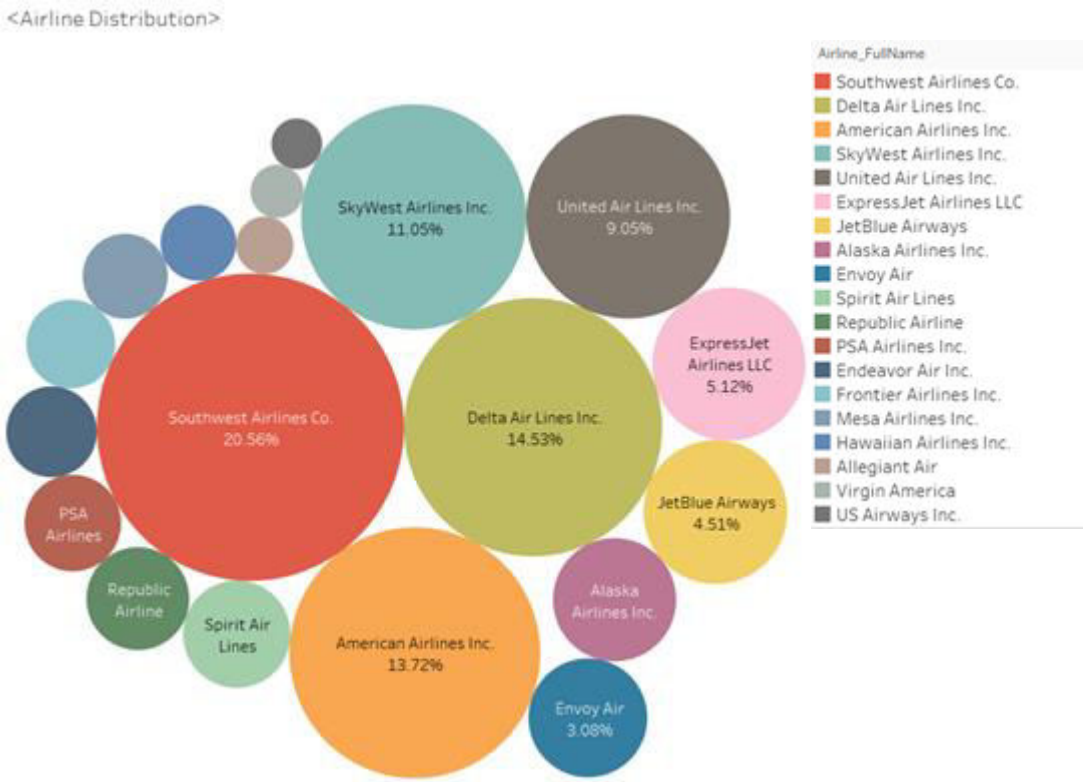
Airline Distribution



5 Dominative Airlines

- Southwest Airlines (20.56%)
- Delta Airlines (14.53%)
- American Airlines (13.72%)
- SkyWest Airlines (11.06%)
- United Airlines (9.05%)

Airline Distribution



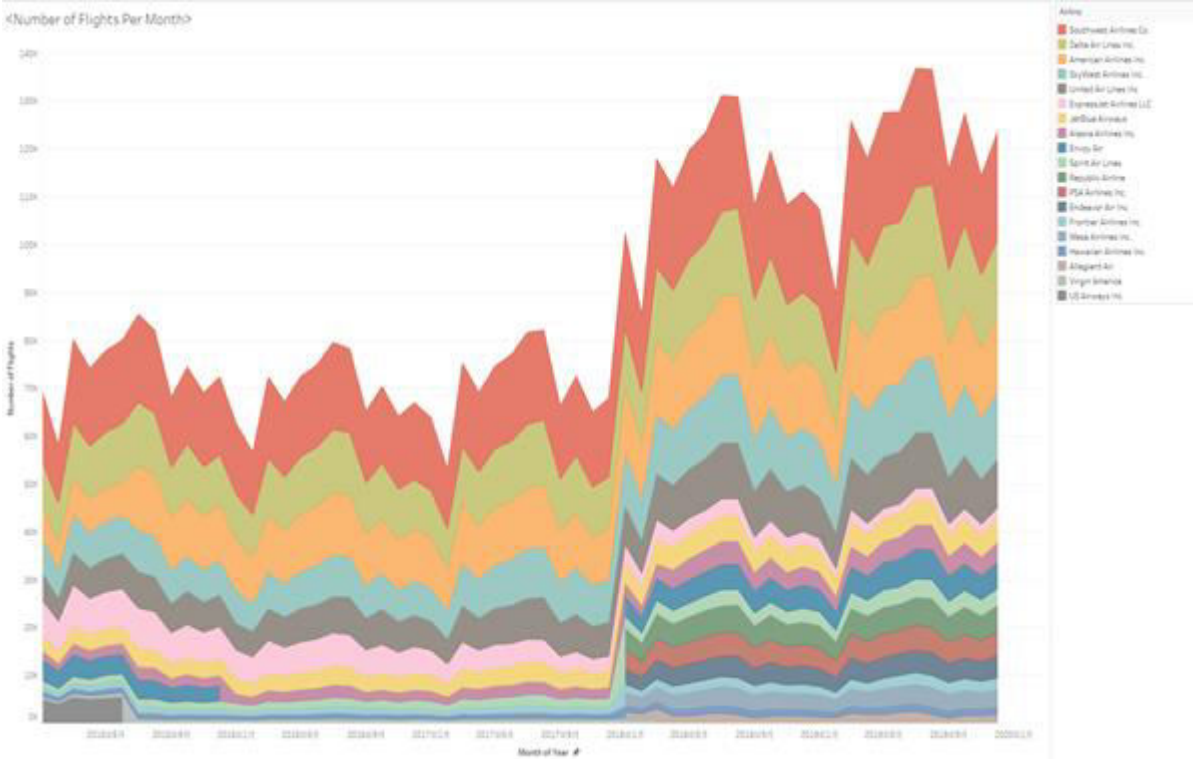
5 Dominative Airlines

- Southwest Airlines (20.56%)
- Delta Airlines (14.53%)
- American Airlines (13.72%)
- SkyWest Airlines (11.06%)
- United Airlines (9.05%)

Total Market Share

~70 %

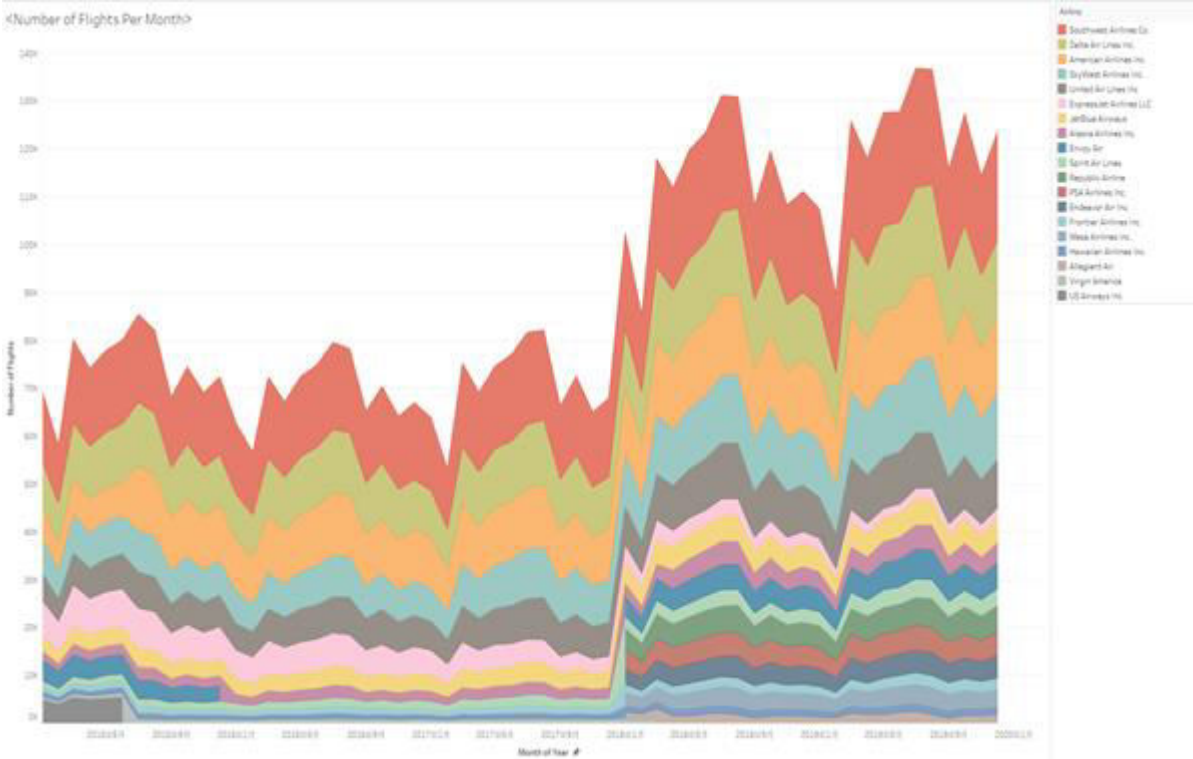
Number of Flights Per Month Throughout 2015-2019



Which airlines have the higher number of flights?

- Southwest Airlines
- Delta Airlines
- American Airlines

Number of Flights Per Month Throughout 2015-2019



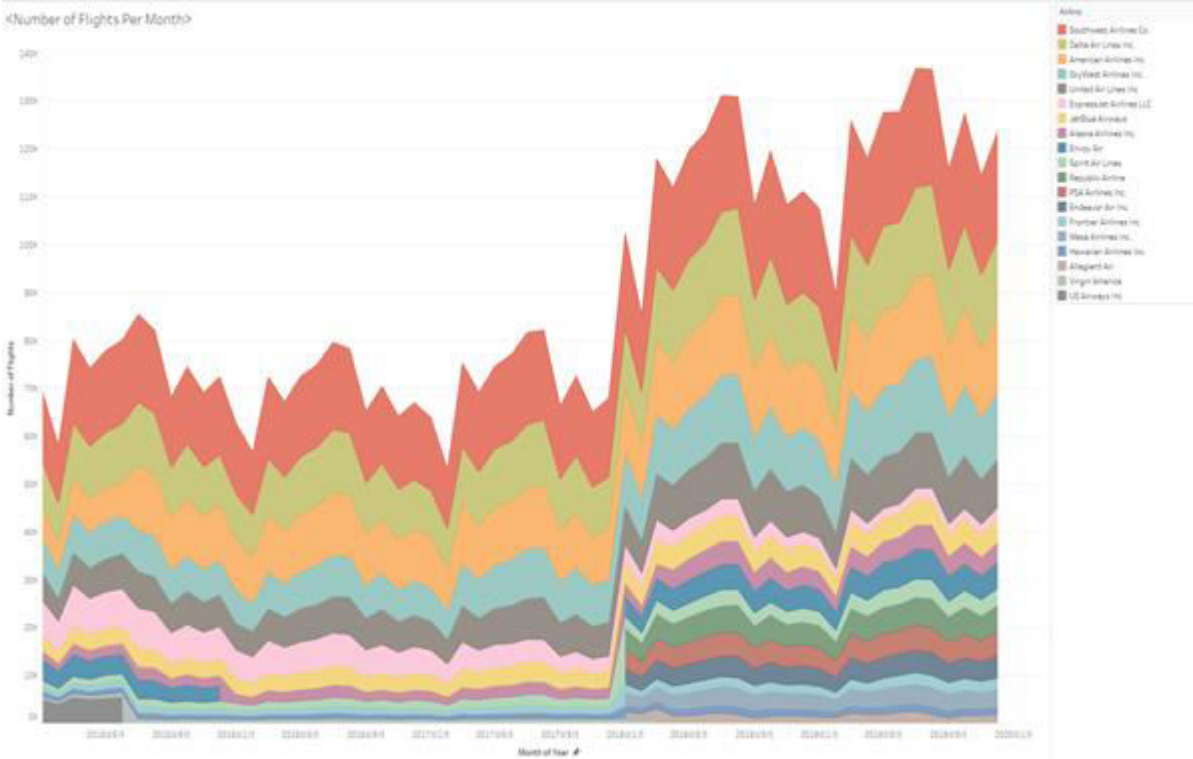
Which airlines have the higher number of flights?

- Southwest Airlines
- Delta Airlines
- American Airlines

Drastic Flights Increase

- From 2017 to 2018
- More airlines enter the market

Number of Flights Per Month Throughout 2015-2019



Which airlines have the higher number of flights?

- Southwest Airlines
- Delta Airlines
- American Airlines

Drastic Flights Increase

- From 2017 to 2018
- More airlines enter the market

Stability in Market Share

- All players were having stable market share (except the change)
- After more airlines entered the market, top players' market share shrinked

Weekday Distribution of Flights by Airlines

Make a guess…

Which weekday has the most frequency of flights?

Weekend VS Weekdays

Weekday Distribution of Flights by Airlines

Make a guess…

Which weekday has the most frequency of flights?

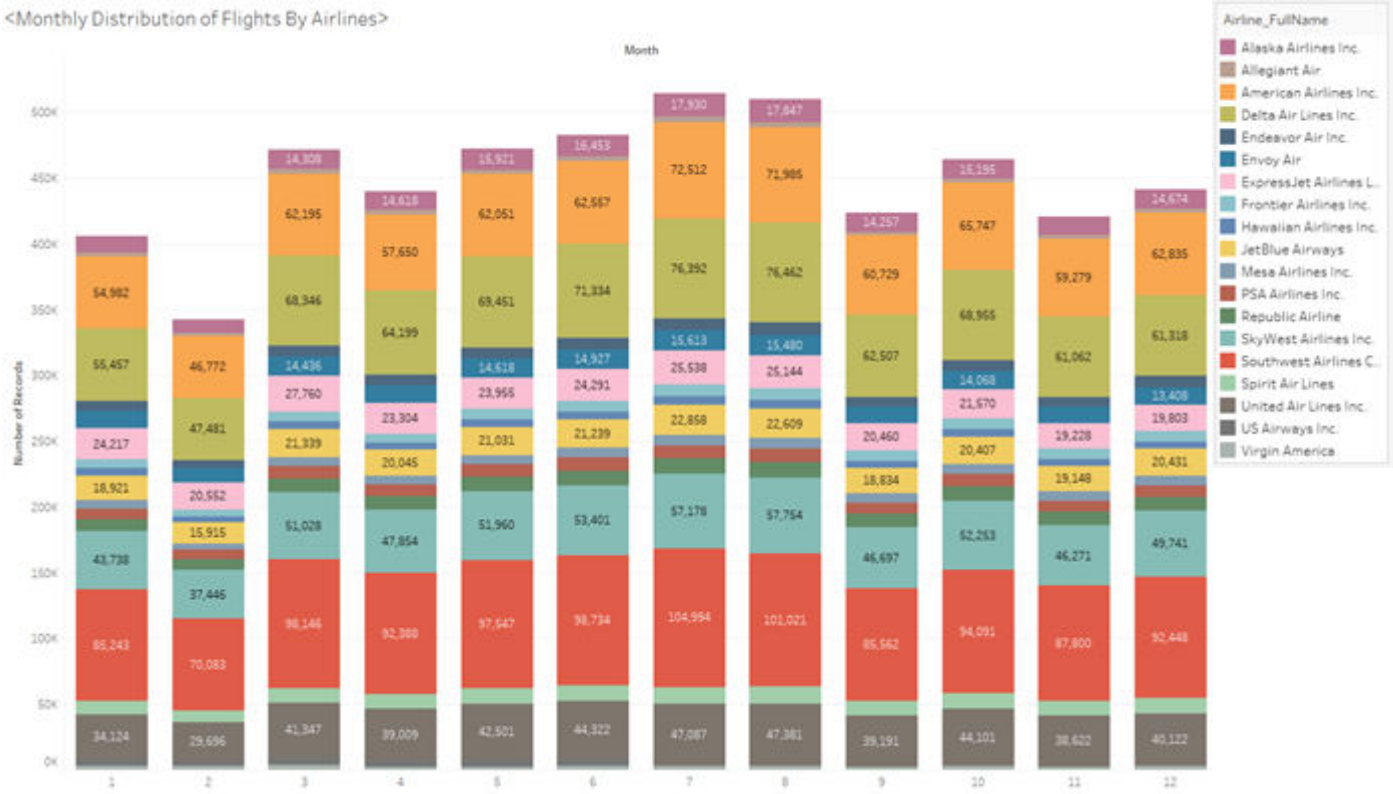
Weekend VS Weekdays

Answer : Weekdays

The top three weekdays are:

I. Friday
II. Monday
III. Thursday

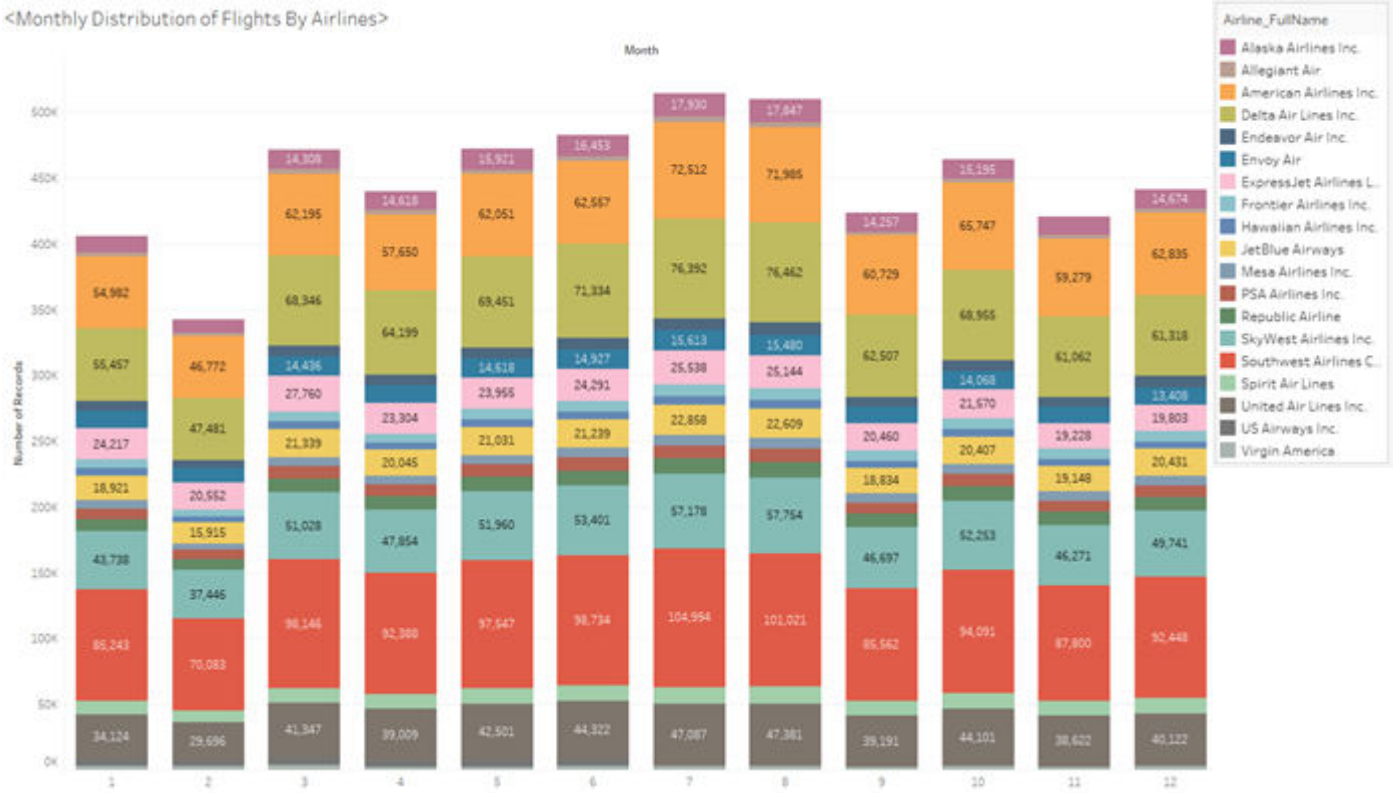
Monthly Distribution of Flights by Airlines



When is the peak season?

- Summer time
- May to August

Monthly Distribution of Flights by Airlines



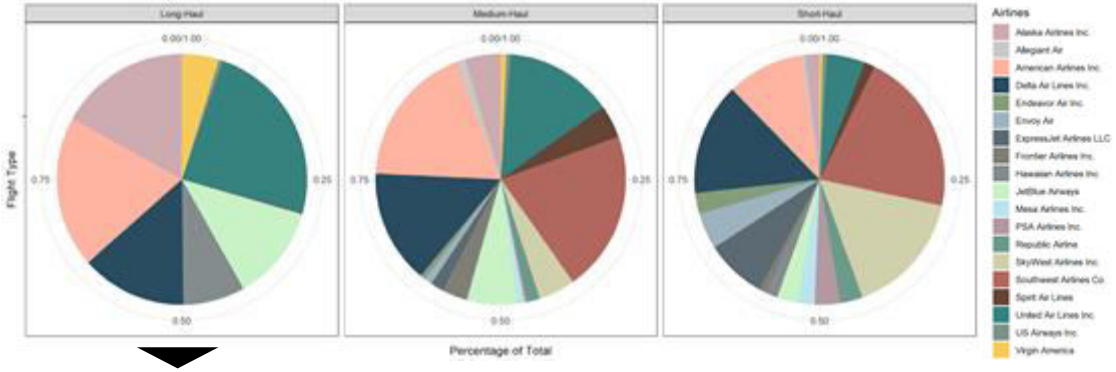
When is the peak season?

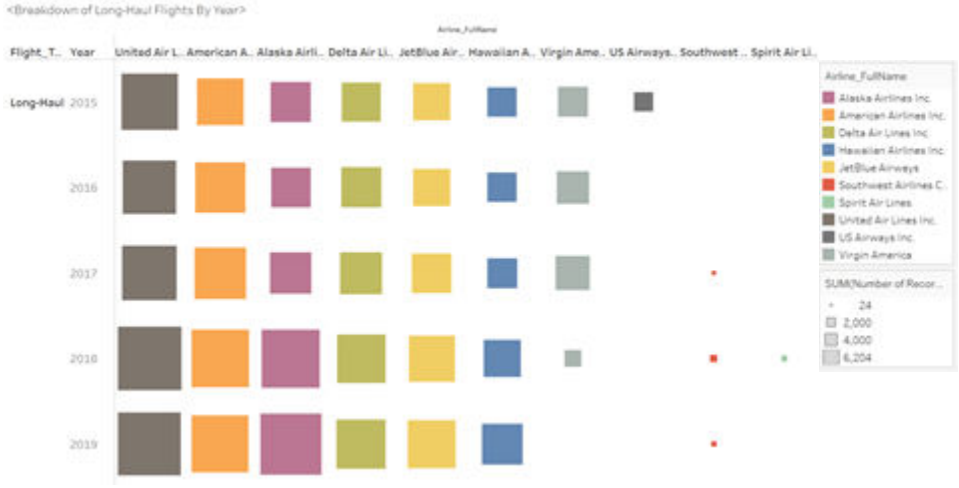
- Summer time
- May to August

Which month is the off season?

- February

Breakdown of Airline Distribution by Haul Types





Offered by least airlines

Long Haul
(>2500 miles)

Medium Haul
(750 miles - 2500 miles)

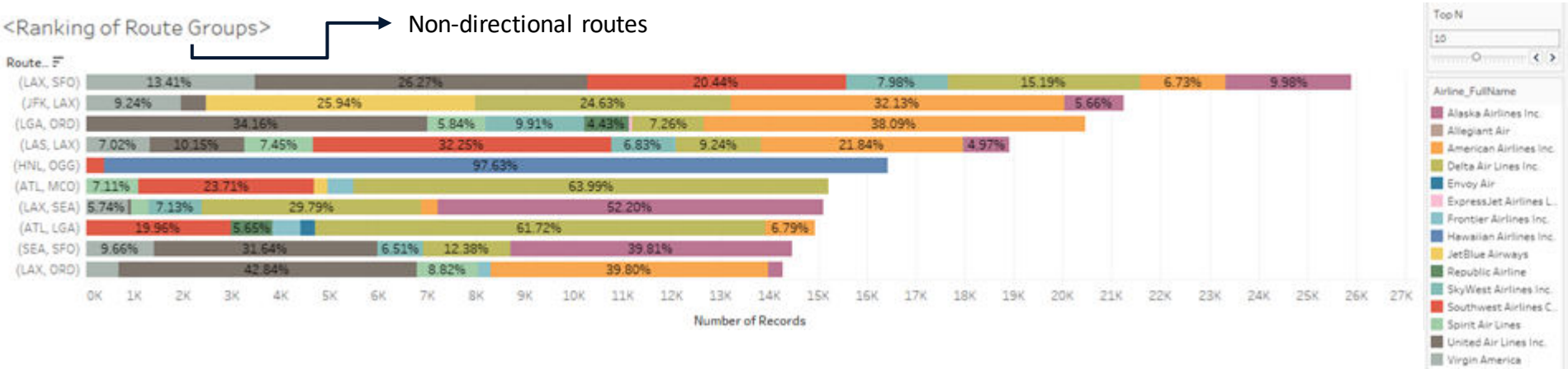
Short Haul
(<750 miles)

Offered by most airlines

Only **10** airlines offer long haul flights:
Alaska Airlines, **American Airlines**, **Delta Airlines**, Hawaiian Airlines, JetBlue Airways, **Southwest Airlines**, Spirit Airlines, **United Airlines**, US Airways, Virgin America

SkyWest Airlines (1 of the top 5 players) does not offer long haul flights

Top 10 Route Groups



Airline Distribution in route groups is not balanced

Common location among these route groups: LAX(5/10), ATL (2/10), SFO(2/10), SEA(2/10)

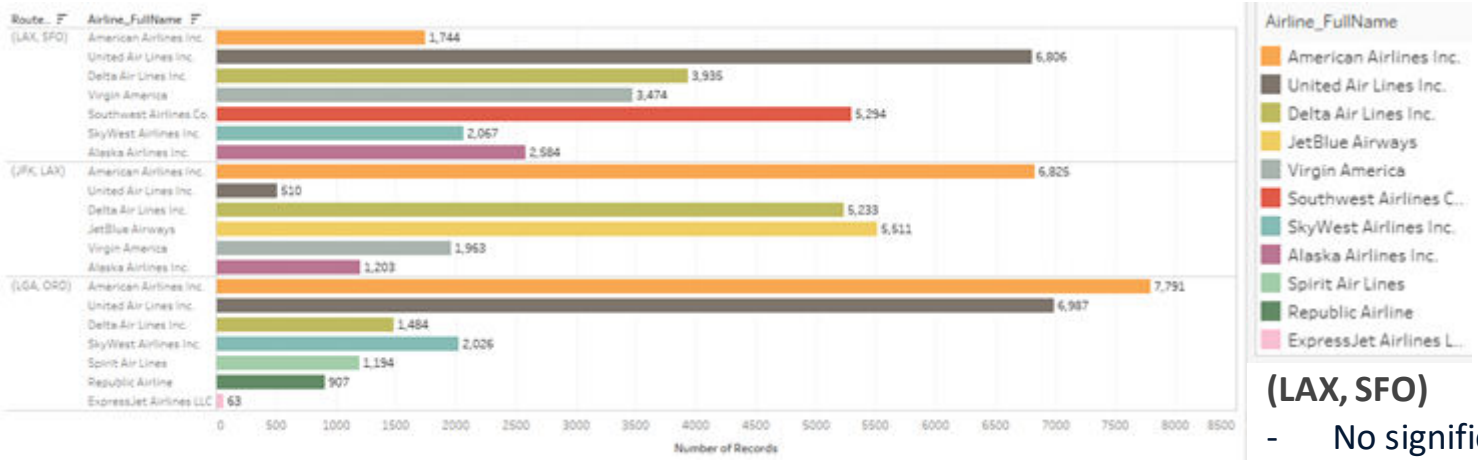
(HNL, OGG): dominated by Hawaiian Airlines *(Hawaii's largest and longest-serving airline)*

(ATL, MCO), (ATL, LGA): dominated by JetBlue Airways

(LAX, SEA): dominated by Alaska Airlines

> 50%

Breakdown of Top 3 Route Groups by Airlines



Only **American Airlines**, **Delta Airlines**, **United Airlines** provide services for all three route groups

ExpressJet Airlines LLC, Republic Airlines, Spirit Airlines only provide service for the (LGA, ORD) route group within these three route groups

(LAX, SFO)

- No significant difference in distribution
- Around 1,000 difference per airline

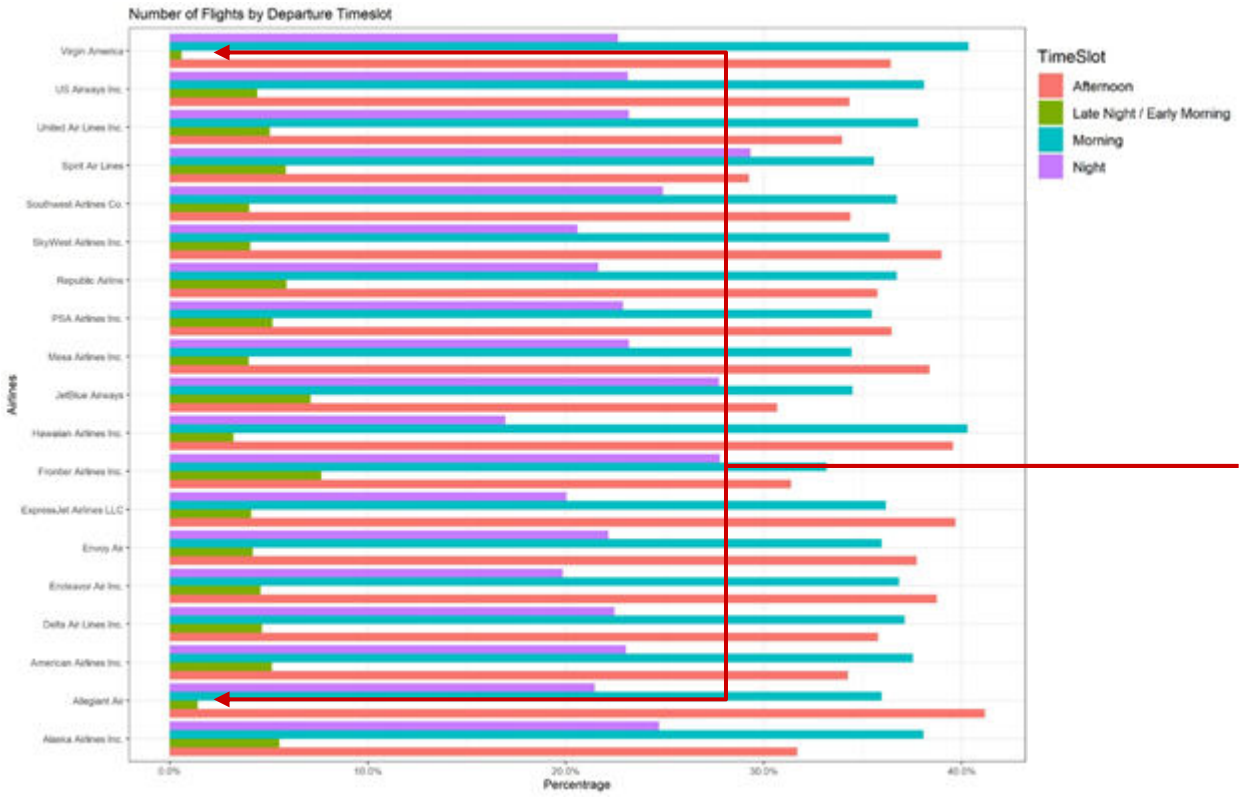
(JFK, LAX)

- 3 Airlines occupied most share
  (American Airlines, JetBlue Airways, Delta Airlines)

(LGA, ORD)

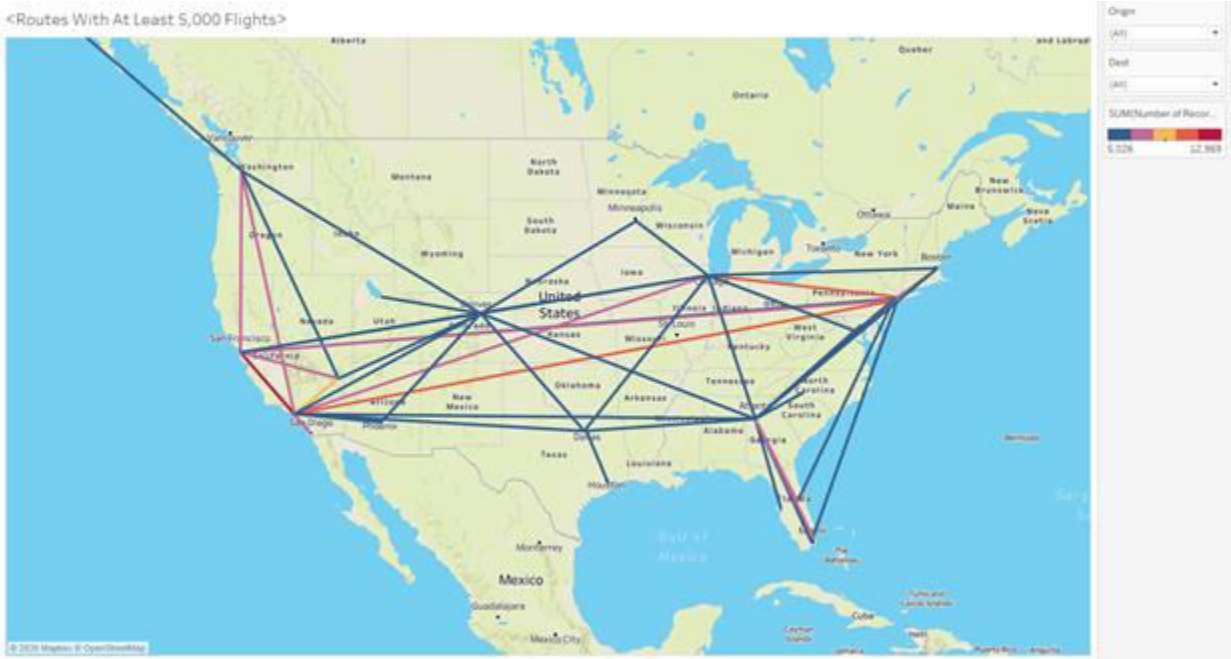
- 2 Airlines occupied most share
  (American Airlines, United Airlines)

Number of Flights By Departure Time by Airlines



- Late Night / Early Morning flights are the least popular
- Afternoon and Morning flights are most popular
- All airlines have similar ratio for the flights' different departure time slots
- Virgin America and Allegiant Air provide much less Late Night / Early Morning flights

Routes With At Least 5,000 Flights



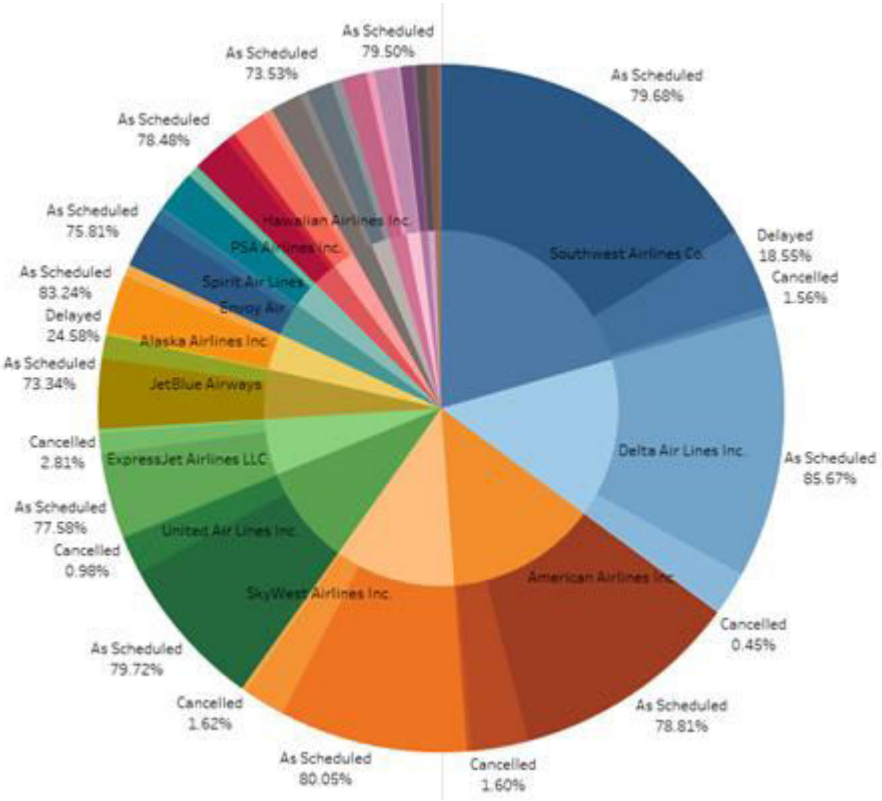
- **Routes slightly above 5,000 flights:**
Locates mostly at the southeast and the southwest regions
- **Busiest airports:**
located at left and right ends & of the central part

Airline Distribution in Top 10 Origins / Destinations

- Same ranking in both origins and destinations
- Delta Airlines has been dominating most flights from / to Atlanta, other airlines serving same location were not able to take over Delta Airlines
- American Airlines was dominating flights from / to Dallas, Charlotte and Phoenix
  In recent years, other airlines are threatening its position in Charlotte and Phoenix

Let’s take a look at each airline’s performance

Distribution of Flight Status by Airlines



Similar distribution of flights status:

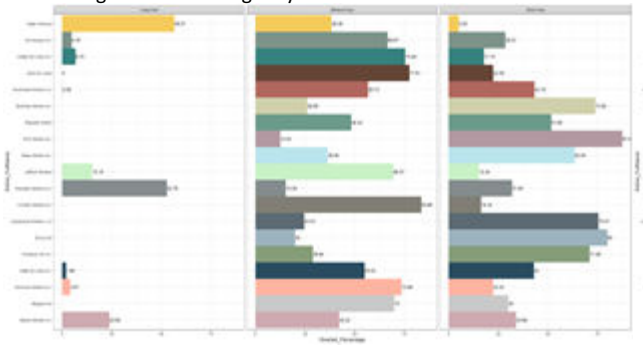
As Scheduled > Delayed > Cancelled > Diverted

Top 5 dominative airlines - As Scheduled

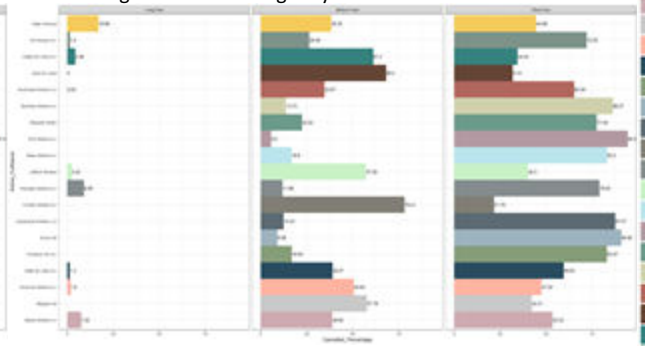
1. Delta Airlines : 85.67%
2. SkyWest Airlines : 80.05%
3. United Airlines : 79.72%
4. Southwest Airlines : 79.63%
5. American Airlines : 78.82%

Breakdown of Percentage of Flights by Haul Type by Each Status

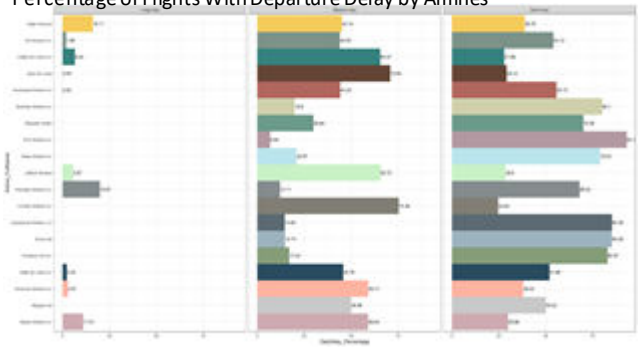
Percentage of Diverted Flight by Airlines



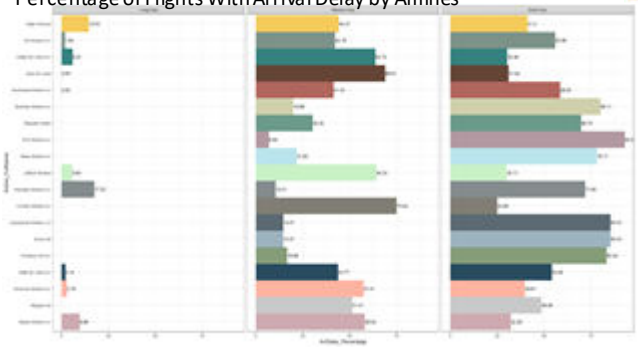
Percentage of Cancelled Flight by Airlines



Percentage of Flights With Departure Delay by Airlines



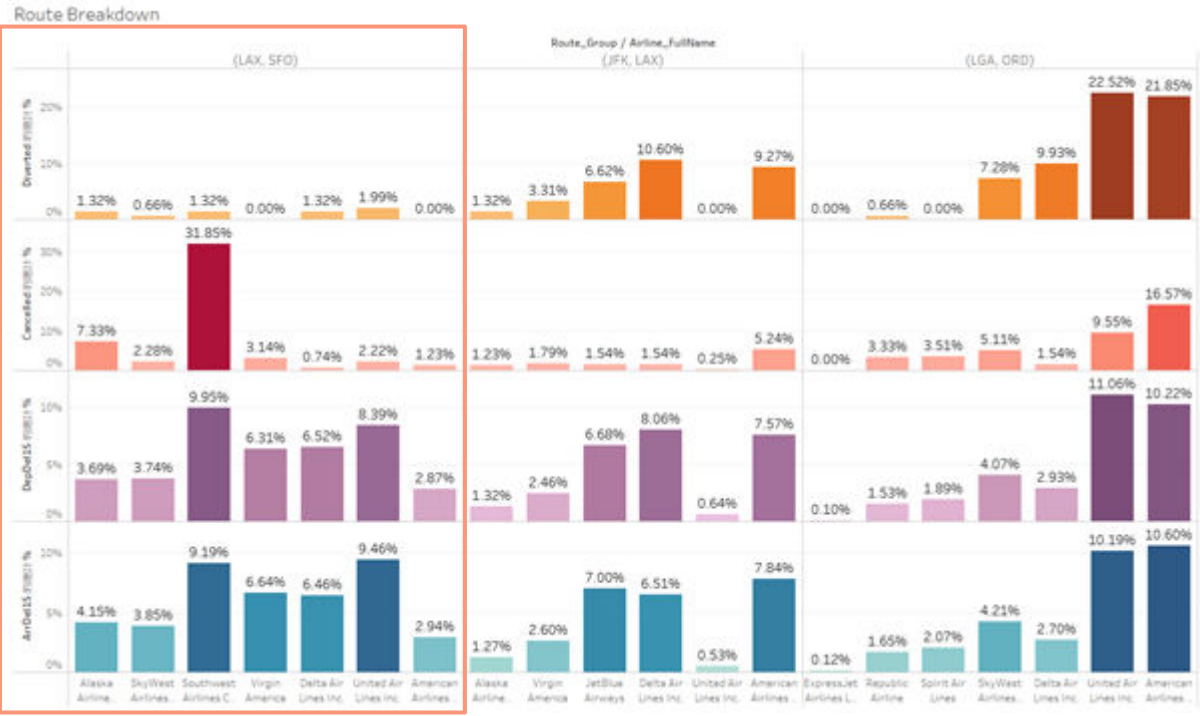
Percentage of Flights With Arrival Delay by Airlines



Top 5 dominative airlines

- **Southwest Airlines & Delta Airlines:**
  a. Ranked middle in all status in short and medium-haul
  b. Ranked near bottom in long-haul
- **American Airlines**
  a. Ranked near bottom in all status in short-haul
  b. Ranked within top 6 in medium-haul
  c. Ranked near bottom in long-haul
- **SkyWest Airlines**
  a. Ranked 2nd highest in all status in short-haul
  b. Ranked near bottom in all status in middle-haul
  c. Ranked near bottom in long-haul
- **United Airlines**
  a. Ranked within bottom 3 in all status in short-haul
  b. Ranked within top 3 in medium-haul
  c. Ranked middle in long-haul

Top 3 Routes Breakdown by Status by Airlines



Tips for Los Angeles (LAX) to San Francisco(SFO)?

Most **Delay**

Southwest Airlines *(~10% Dep Delay, ~9.2% Arr delay)*

United Airlines *(~8.4% Dep Delay, ~9.5% Arr delay)*

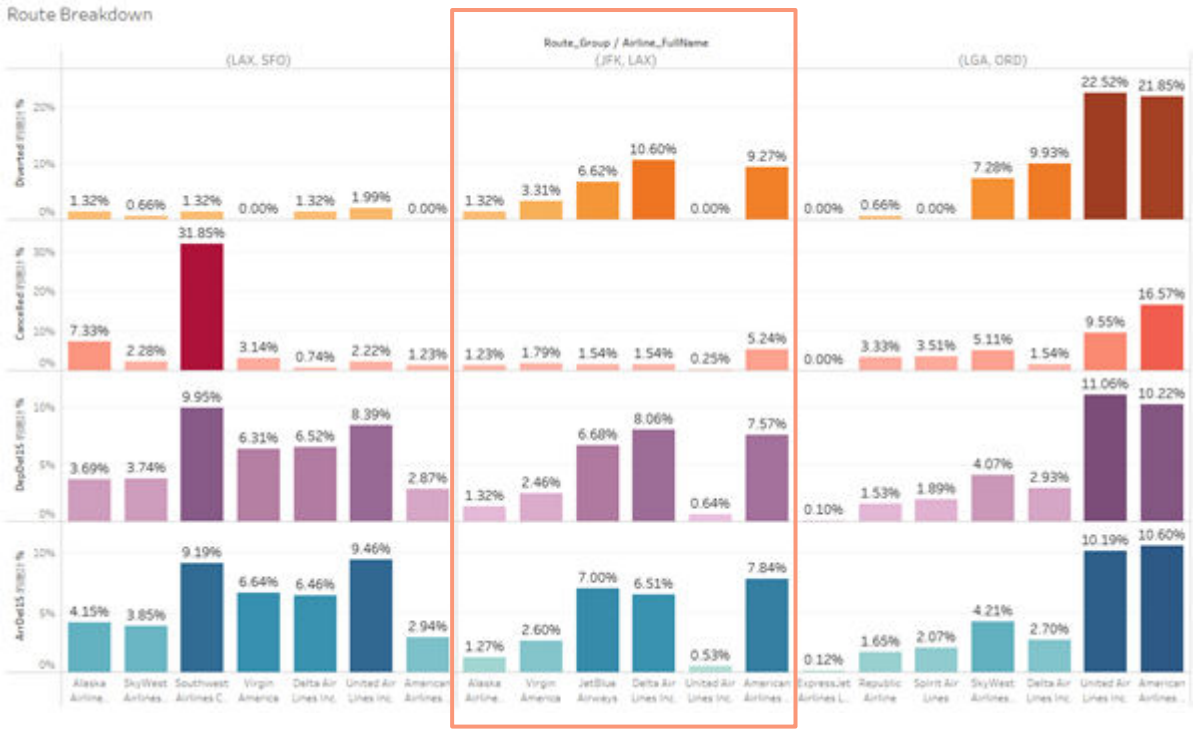
Virgin & Delta Airlines *(~6.5% Dep & Arr delay)*

Most **Cancellation** for Southwest Airlines

(~31.8%)

Southwest Airlines should be avoided if the schedule is tight

Top 3 Routes Breakdown by Status by Airlines



Tips for New York (JFK) To Los Angeles(LAX)?

Most **Delay**

American Airlines *(~7.6% Dep delay, ~ 7.8% Arr delay)*

JetBlue Airways *(~6.7% Dep delay, ~7% Arr delay)*

Delta Airlines *(~8% Dep delay, ~6.5% Arr delay)*

Most **Cancellation** for American Airlines *(~5.2%)*

Most **Diverted**

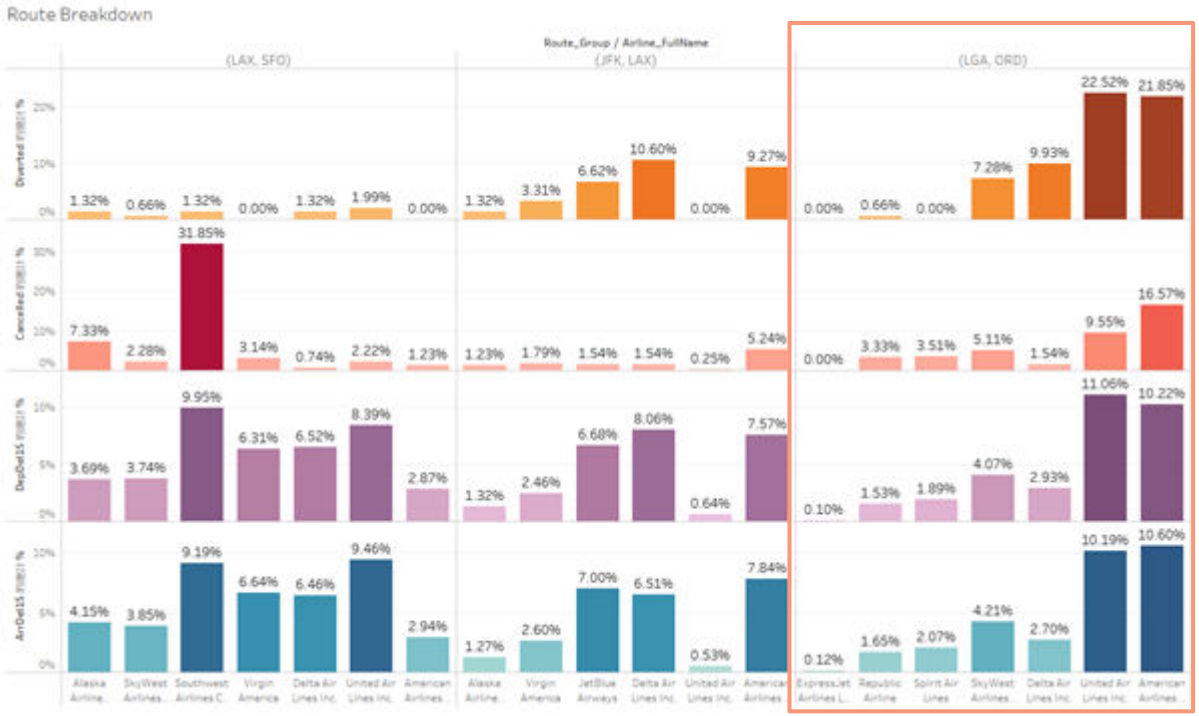
Delta Airlines *(~10.6%)*

American Airlines *(~9.3%)*

JetBlue Airways *(~6.6%)*

American Airlines, JetBlue Airways & Delta Airlines should be avoided if the schedule is tight

Top 3 Routes Breakdown by Status by Airlines



Tips for New York (JFK) to Chicago(ORD)?

Most **Delay**

American Airlines *(~10.2% Dep Delay, ~10.6% Arr Delay)*

United Airlines *(~11.1% Dep Delay, ~10.2% Arr delay)*

Most **Cancellation** for American Airlines *(~16.57%)*

Most **Diverted**

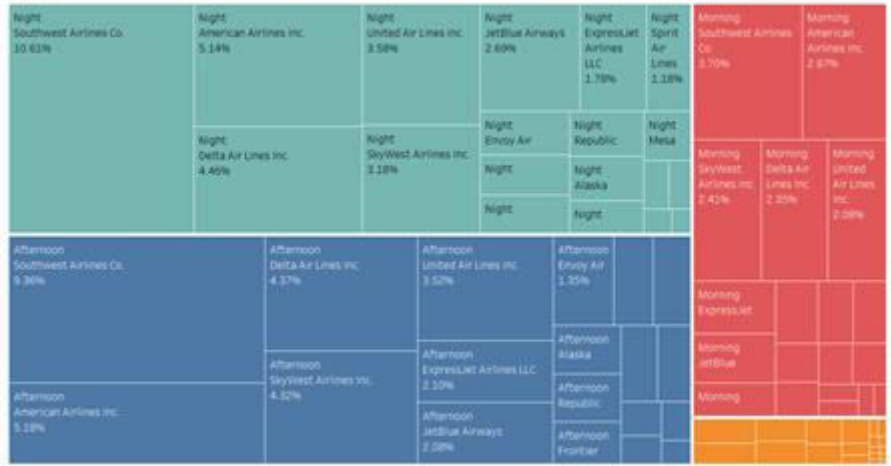
United Airlines *(~22.52%)*

American Airlines *(~21.85%)*

Delta Airlines *(~9.93%)*

American Airlines, United Airlines should be avoided if the schedule is tight

Departure and Arrival Delay by Airlines



Airline Most Frequent Airline in Departure & Arrival Delay:

- Southwest Airlines, in all time slots tops all

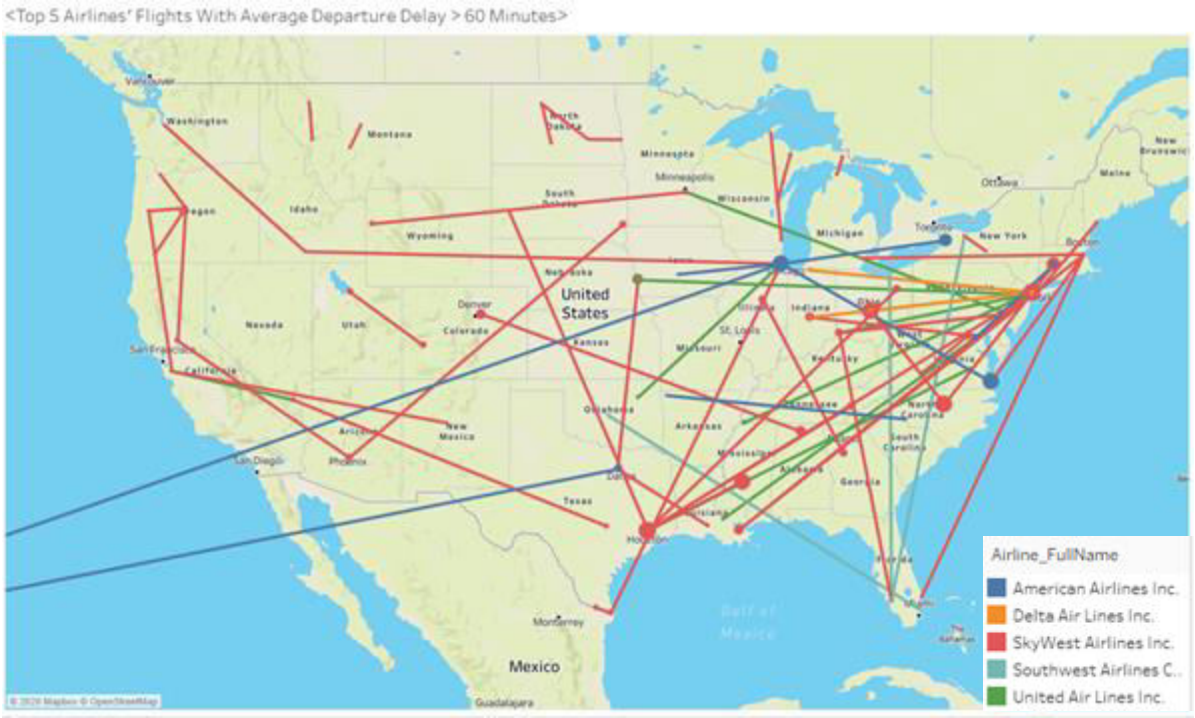
However, it manages to arrive on time in some late departures

- Southwest Airlines: Night late departure: 10.61% → Night late arrival: 10.44%
- Southwest Airlines: Afternoon late departure: 9.3% → Afternoon late arrival: 6.13%

In Comparison,
American Airlines, Delta Airlines, United Airlines, SkyWest Airlines

❖ Sometimes **arrive late** even don't have late departure at night

Top 5 Airlines' Flights With Average Departure Delay > 60 Minutes



1st: SkyWest Airlines

- Red lines show that a serious departure delay problem especially in the Northeast

2nd: United Airlines

- Green lines show that a high tendency of short haul flights departure delay, especially in the Northeast

3rd: American Airlines

- Blue lines show that an often long haul flight departure delay, especially from Chicago and Richmond

Top 5 Airlines' Flights With Average Arrival Delay > 60 Minutes

You may think serious departure delay must cause a serious arrival delay, but that's not true

- A different graph between departure delay & arrival delay

Routes that have no departure delay but arrival delay > 60 mins

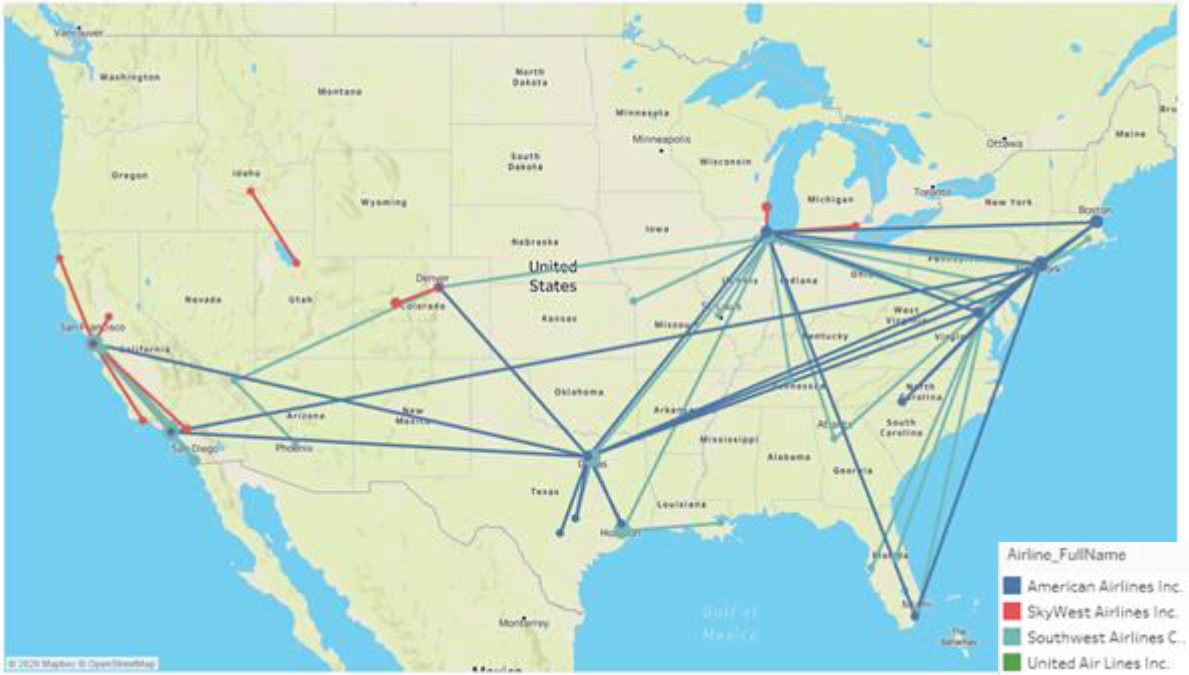
- **Delta Airlines:**
  Topeka - Las Vegas,
  Las Vegas - Vancouver*
- **United Airlines:**
  Denver - Fresno

*Vancouver (City in Washington)

Routes that have departure delay but no arrival delay > 60 mins

- **American Airlines:**
  Long Haul Flights from Chicago
- **SkyWest Airlines:**
  San Francisco - Phoenix

Top 5 Airlines' Flights With Cancellation > 50



1st: American Airlines

- Mainly in the Northeast
- New York, Dallas and Chicago are serious spots

2nd: Southwest Airlines

- Mainly in the Northeast
- San Francisco, San Diego and Dallas are serious spots

Top 5 Airlines' Flights With Diversion > 20

Route

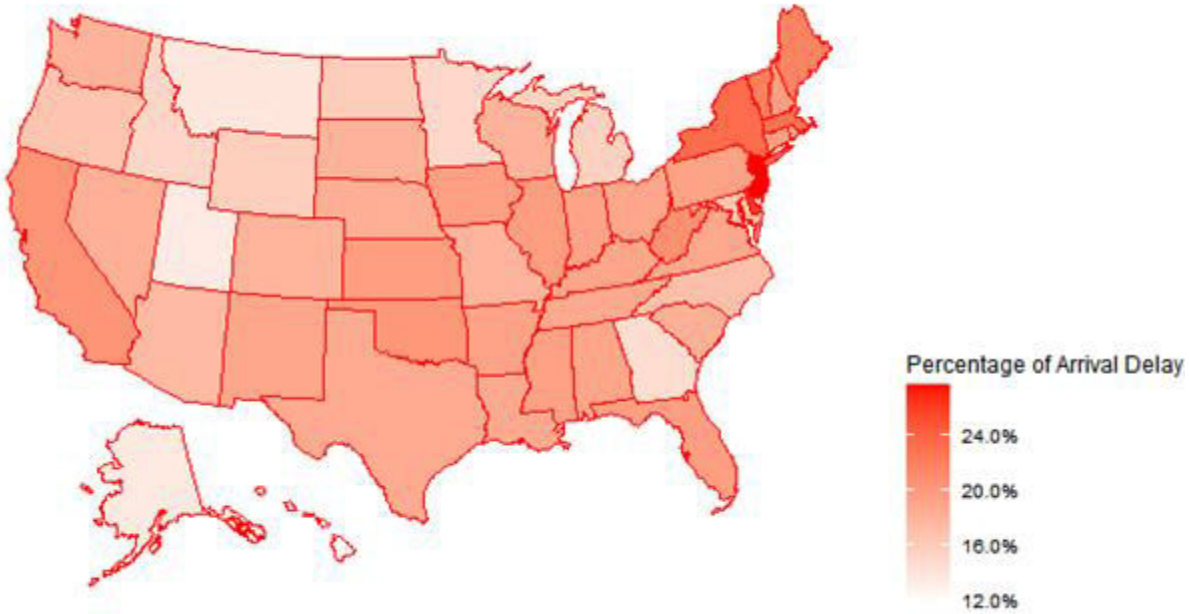
- **Long-haul flights** tend to have more diversions

Airlines

- **SkyWest Airlines** tend to have more diversions

Let’s take a deeper look at Delay

The Percentage of Arrival Delay by Destination States



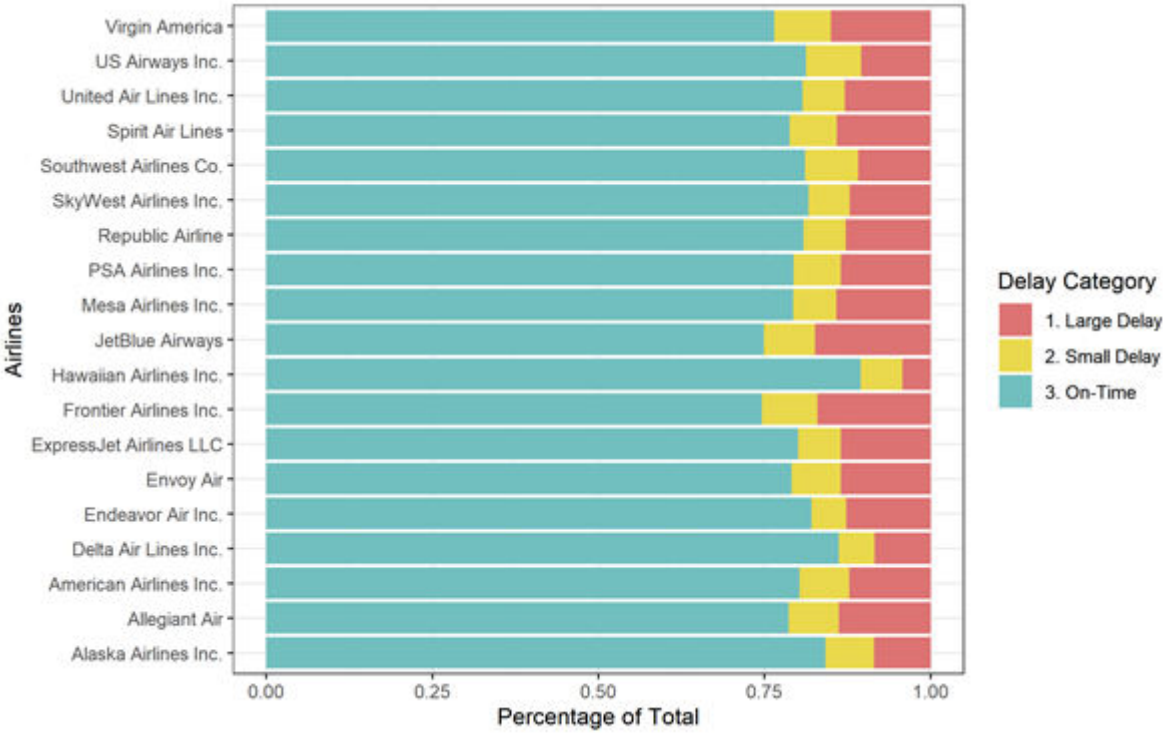
What destinations have the highest percentage of arrival delay?

- Northeast: The northeast States
  - Especially New Jersey, New York Massachusetts & Vermont
- West: California

What destinations have the least percentage of arrival delay?

- Midwest as a belt

Arrival Delay by Delay Type



In general, which airline is mostly on time?

Hawaiian Airlines

What airlines have the most large delays?

Most Frequent
Large Delays

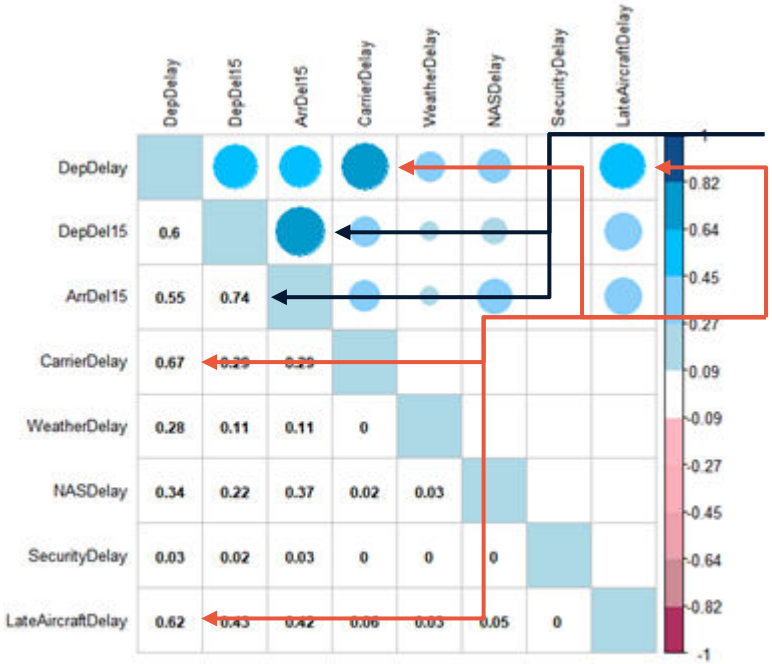
JetBlue Airways, Frontier Airlines

Virgin America, Spirit Airlines, Allegiant Air

Top 5 dominative airlines ranking:

1. Delta Airlines
2. SkyWest Airlines
3. Southwest Airlines
4. American Airlines
5. United Airlines

Correlation Matrix Between Variables

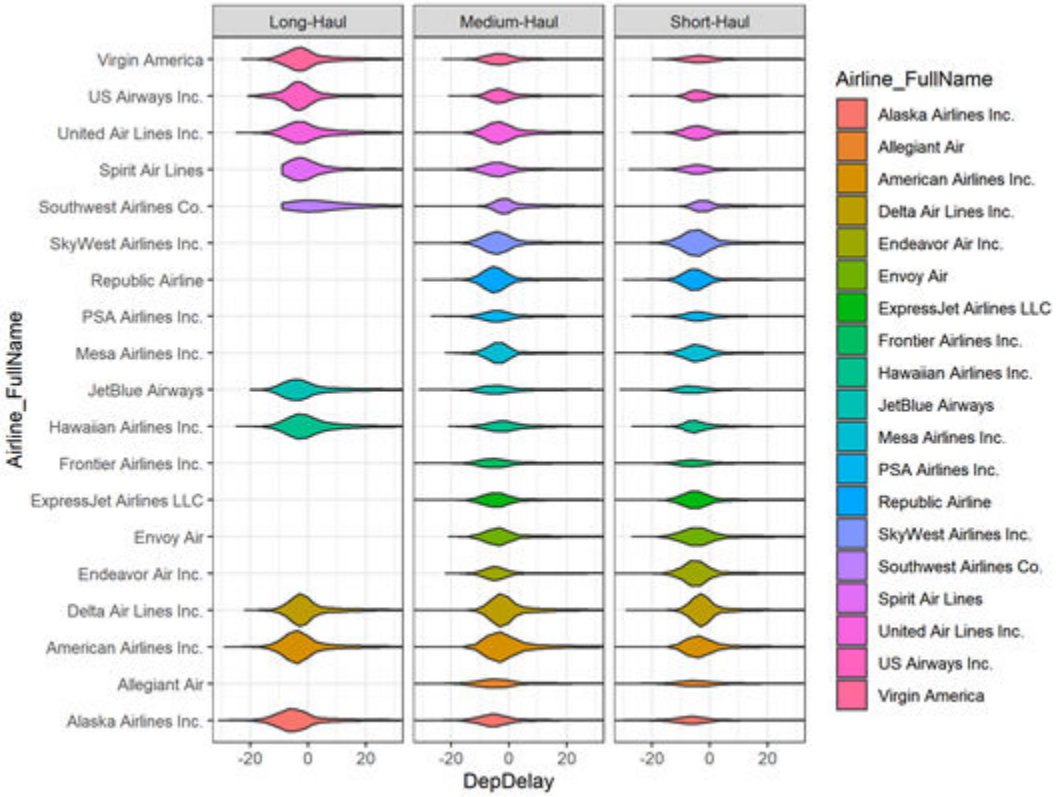


Here we see several significant correlations:

(1) There is a **large** positive linear association between Departure Delay and Arrival Delay (0.74)

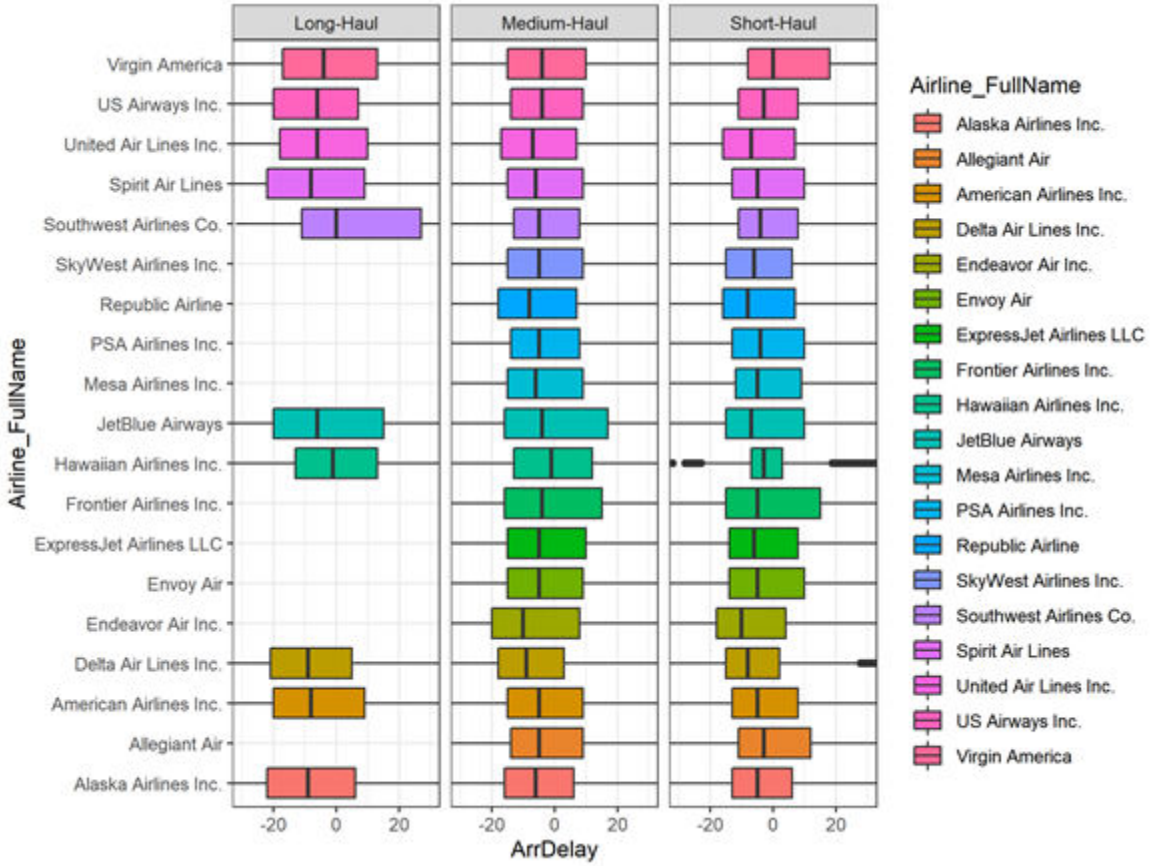
(2) There is a **moderate** linear relationship between Departure Delay and Carrier Delay (0.67), and Late Aircraft Delay (0.62)

Distribution of Departure Delay from -30 minutes to 30 minutes by Flight Types



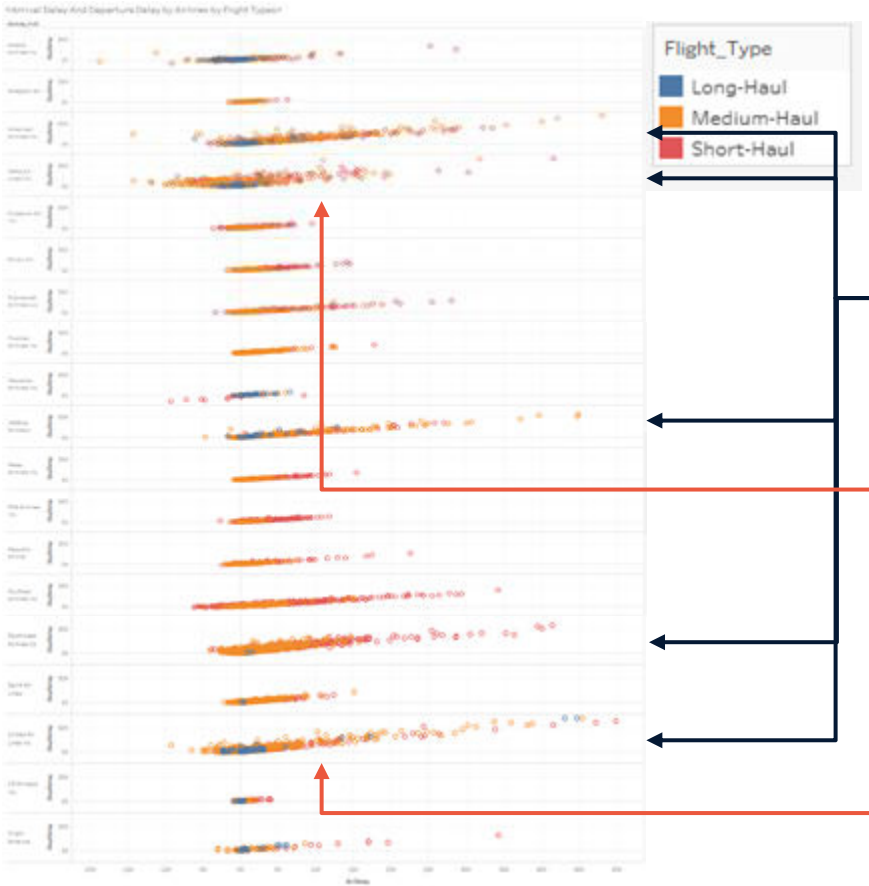
- Longer the haul is, higher distribution of departure delay.
- Most of the distribution stays at -10 minutes and tend to be left hand sided.

Distribution of Arrival Delay from -30 minutes to 30 minutes by Flight Types



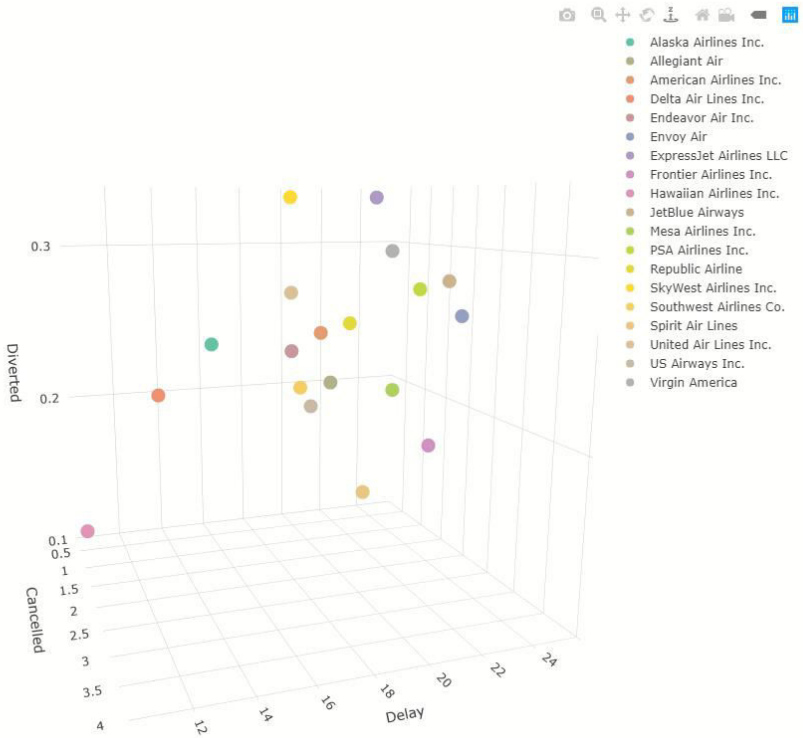
- Long-Haul has a wider range of distribution
- Hawaiian Airlines adheres to scheduled arrival time
- Median of distribution mostly lay on -10 minutes
- Most distribution are left-skewed
- Frontier Airlines and JetBlue Airlines has the largest upper-quartile

Average Arrival Delay vs Average Departure Delay by Airlines



- Within 3 hauls, **long-haul** tends not to have arrival/departure delay
- American Airlines, Delta Airlines, Southwest Airlines, JetBlue Airways and United Airlines tend to have more severe delays
- Delta Airlines and United Airlines were able to achieve more negative delays

Percentage of Flights with Arrival Delay, Cancellation & Diversion by Airlines



Best 3 (Arrival Delay, Cancellation, Diversion):

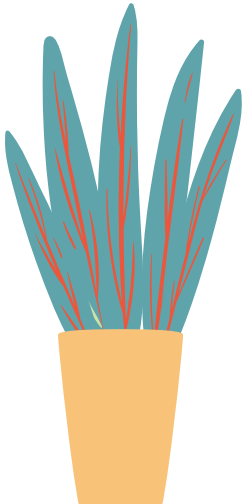
1. Hawaiian Airlines (10.5, 0.2, 0.1)
2. Delta Airlines (13.7, 0.5, 0.2)
3. Alaska Airlines (15.8, 0.8, 0.2)

Performance of the top 5 market players:

- Southwest Airlines (18.9, 1.6, 0.2)
- Delta Airlines (13.7, 0.5, 0.2)
- American Airlines (19.7, 1.6, 0.2)
- SkyWest Airlines (18.4, 1.6, 0.3)
- United Airlines (19.3, 1.0, 0.3)

05

Summary



Summary

- There are 5 airlines dominating the aviation industry in US throughout 2015-2019
  1. Southwest Airlines
  2. Delta Airlines
  3. American Airlines
  4. SkyWest Airlines
  5. United Airlines
- Summer is the peak season
- Some airlines particularly serve a specific market / some specific markets (e.g. Hawaiian Airlines)
- Late night / Early Morning flights are least served
- Southeast & West regions are more frequently travelled to
- Airlines are expanding routes to serve throughout the years, especially for smaller airlines
- Airlines typically have 70-80% flights as scheduled
- Performance of biggest 5 players in the market vary a lot
- (LAX, SFO) Southwest Airlines should be avoided if schedule is tight
- (JFK, LAX) American Airlines, JetBlue Airways & Delta Airlines should be avoided if tight schedule
- (LGA, ORD) Should avoid American Airlines, United Airlines if tight schedule
- Southwest Airlines has most amount of departure delay & arrival delay
- American Airlines has highest cancellation rate
- SkyWest Airlines has highest diversion rate

06

Future Directions

Future Directions

1. Analyze how airlines can improve their performances
2. Analyze the COVID-19 effects on flights
3. Analyze the Hong Kong International Airport and the airlines headquartered in Hong Kong

THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.