TRAVEL AIRLINE DELAY ANALYSIS

IMPACT ON ARRIVAL FLIGHTS TRAVEL ACROSS THE US 2018-2022

Rutgers Data Analytics and Visualization Bootcamp

FEB 2023 TEAM 3: CALOGERA MCCORMICK, TAWN SCOTTON, JAMIE CID, SANA AYUBZAI, MINTA BURKE

HYPOTHESIS

- Recent weather changes have caused an **increase in the number of delays** of arrival flights in **all regions** across the US.

- The **West** and **East Coast** regions have experienced the highest increase in delayed flights.

- Of the top 5 airlines in the US, **American Airlines** experienced the most delays, cancellations and diversions.

Rationale for Analysis

- **Timeframe** – 5 years 2018-2022 provides the same duration of analysis time before and after COVID

- **Airlines** – based on size of fleet in US on 2022 values (or most recent available)

- **Airports** – based on at least 1 month >= 50 weather related delays from at least 1 airline in 2018 to 2022 (18 airports))

- **Regions** – analyze based on 4 regions of the US per the US census (North, South, Midwest and West)

Analysis Process

- **Data collection**: publicly available dataset from the US Bureau of Transportation Statistics (https://www.bts.gov)

- **Data analysis**: Pandas, Matplotlib, Excel

- **Data visualization**: Matplotlib, Excel

RATIONALE FOR HOW AIRLINES WERE CHOSEN

Largest fleet size in US

	Fleet size
American Airlines	1440
United	1391
Delta	1170
Southwest	737
SkyWest	509
Air Canada	353
Alaska Airlines	329
JetBlue	270
West Jet	172
Spirit Airlines	157

Most number of passengers carried per year

	Millions of passengers
American Airlines*	200
Delta*	200
Southwest*	160
United*	140
Air Canada (Canada)	13
Alaska Airlines	40
JetBlue	14
Spirit Airlines	19
WestJet	25
Volaris (Mexico)	24

Variables Researched

Change in number of delayed flights per year per airline

Change in number of delayed flights per year per airport

Change in number of delayed flights due to weather per year per airline

Compare number of delayed flights per year per region within North America (Northeast, South, Midwest, West)

Changes to number of overall delays, cancellations and diversions

Change in number of delayed flights due to weather per year airport

Source: On-Time Data | Bureau of Transportation Statistics (bts.gov)

Definitions

- Arrived – flight arriving at airport

- Delayed – flight arriving more than 15 minutes late

- Cancelled – flight cancelled

- Diverted – flight sent to another destination

- **Delay Reasons**
    - **Airline** – Number of flights delayed due to air carrier. (e.g. no crew)
    - **Weather -** Number of flights due to extreme weather.
    - **NAS** – Number of flights delayed due to National Aviation System (e.g. heavy air traffic).
    - **Security** – Number of flights canceled due to a security breach.
    - **Late aircraft** – Number of flights delayed as a result of another flight on the same aircraft delayed

Overview of Flight Delays (2003-2022)

Number of Flights Delayed, Cancelled, or Diverted in US for Any Reason (2003-2022)
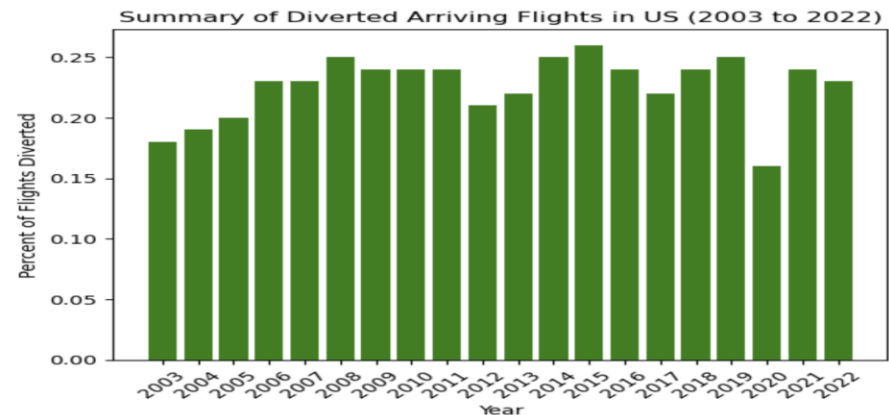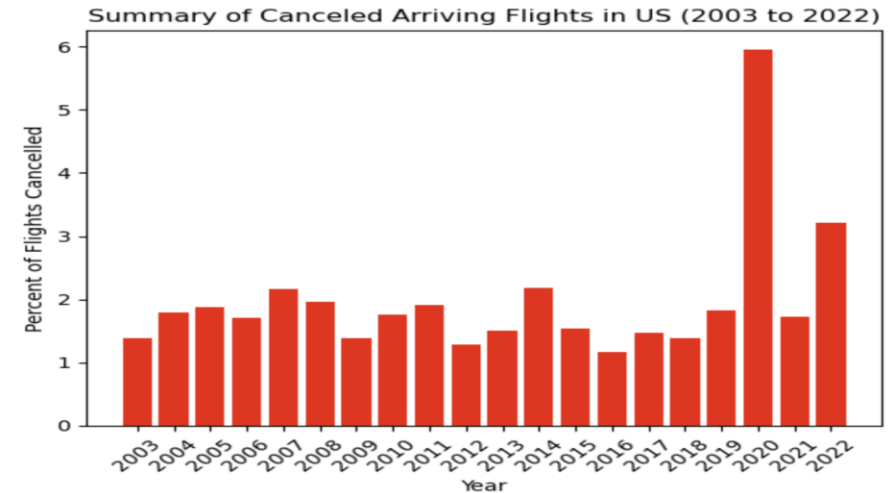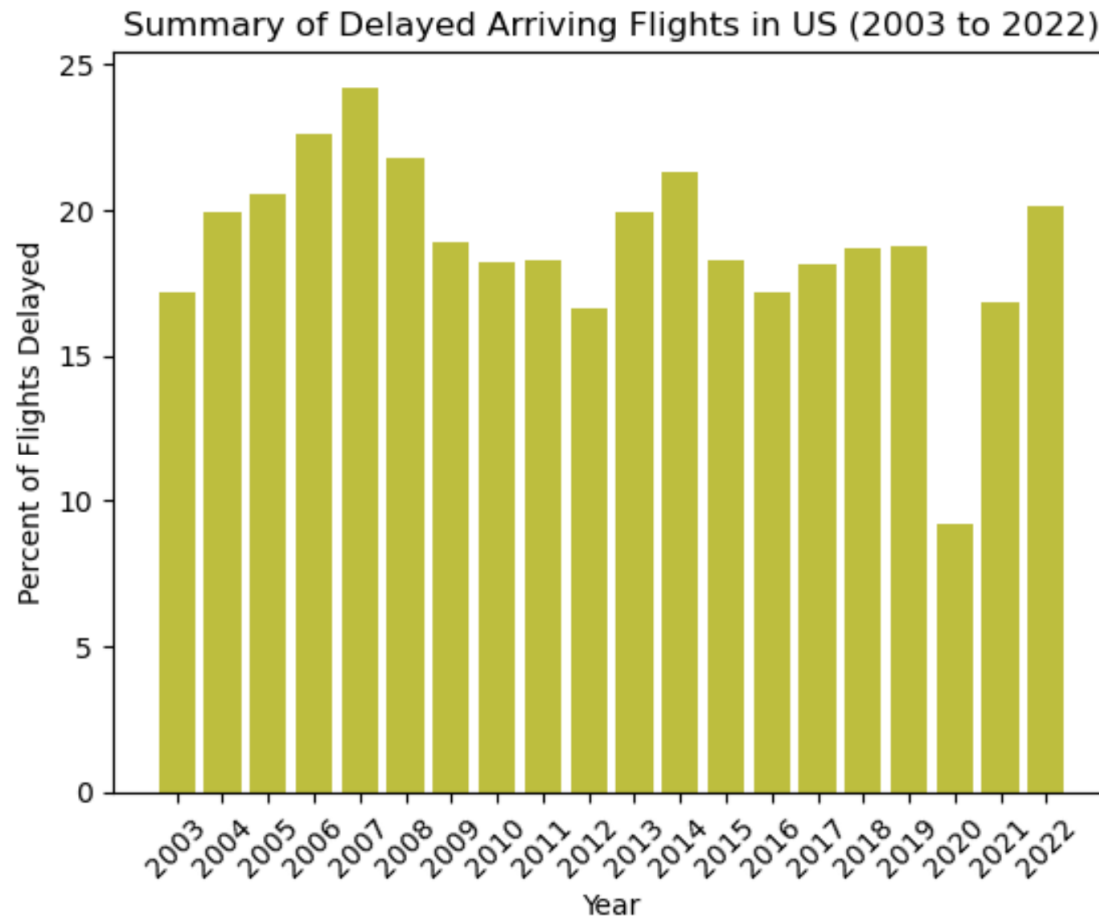
Number of (%) of Flights Delayed, Cancelled, or Diverted in US for Any Reason (2018-2022)

- Relative to the number of arriving flights each year (all airlines and airports aggregated for US), generally more flights are delayed (17-20%) compared to cancelled or diverted.

Year	Arrived	Delayed	Cancelled	Diverted	% Delayed	% Cancelled	% Diverted
2018	6304506.0	1179152.0	87811.0	15265.0	18.70	1.39	0.24
2019	7422029.0	1389253.0	134918.0	18879.0	18.72	1.82	0.25
2020	4685808.0	431871.0	279063.0	7730.0	9.22	5.96	0.16
2021	5995392.0	1010332.0	103055.0	14117.0	16.85	1.72	0.24
2022	2733118.0	549829.0	88134.0	6169.0	20.12	3.22	0.23

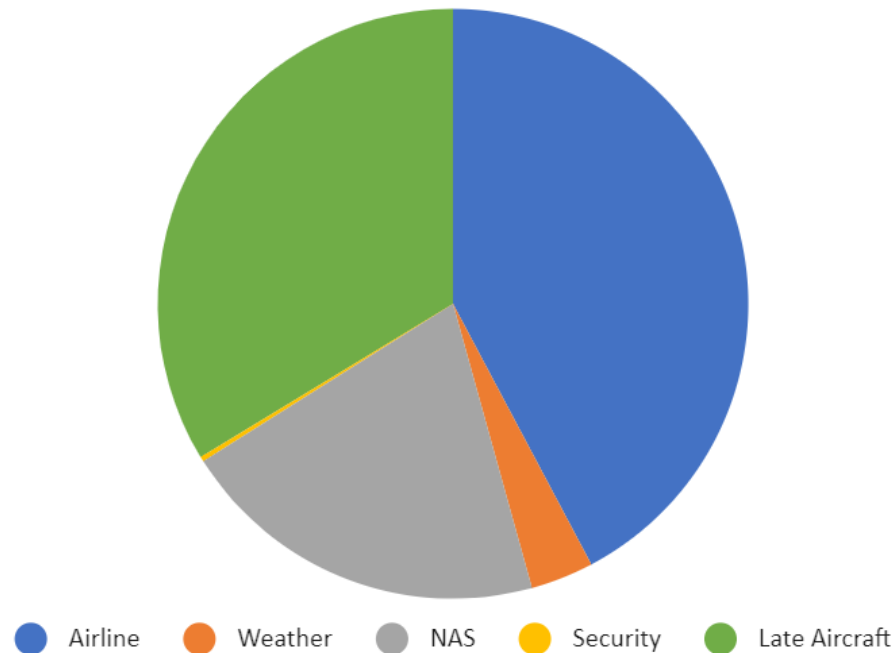# Number of Flight Delays for 18 Airports (2018 to 2022)
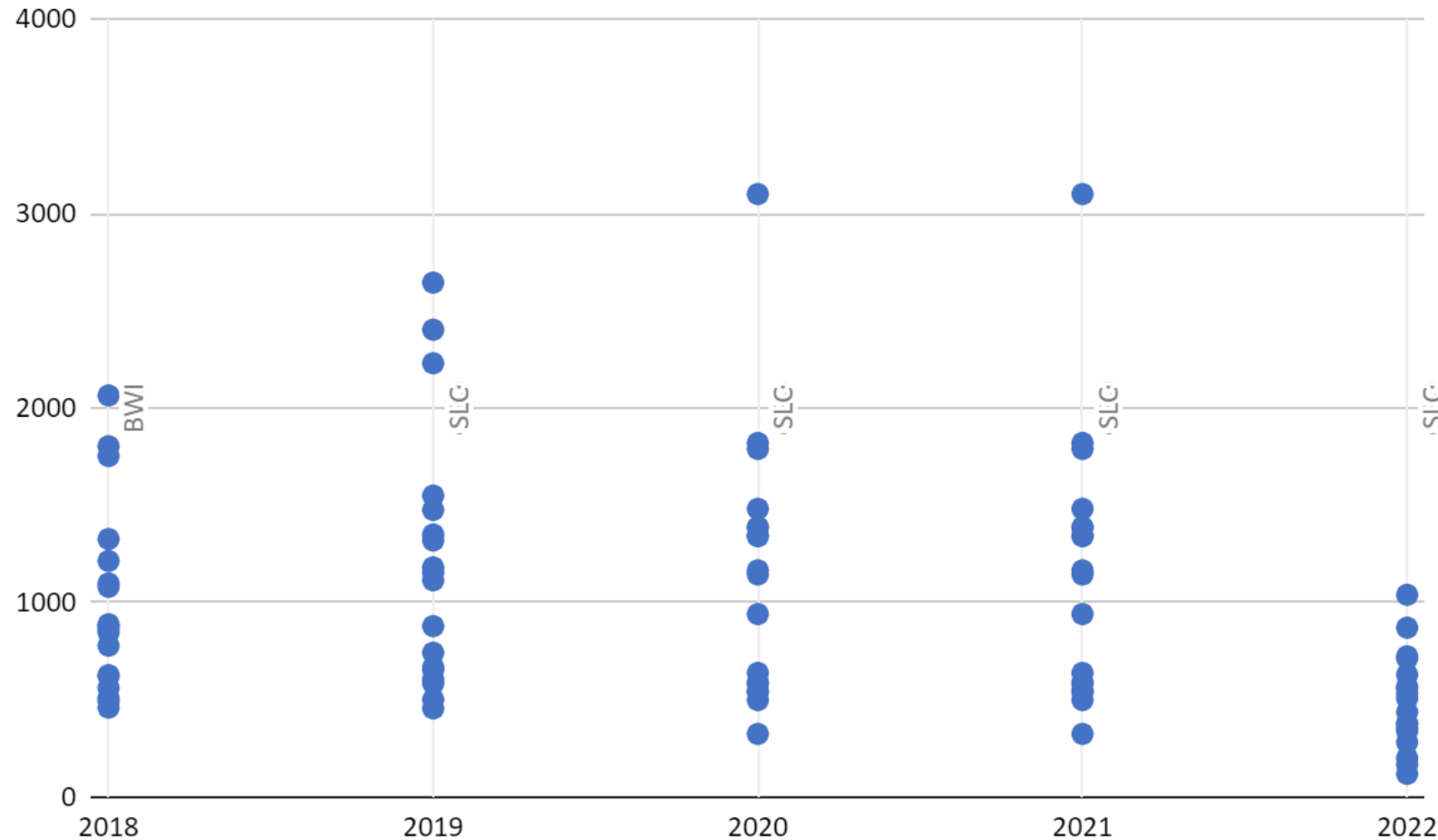
Outliers

South
Dallas - over 2500 in 2019
Dallas - over 3000 in 2020 and 2021
Atlanta - over 2000 in 2018 and 2019

Midwest
Chicago - ~2,400 in 2019

Number (%) of Delayed Flights as % of Total Delayed Flights Per Reason for Delay in US (2022)

- Based on the information given, there were a total of 333,916 delayed flights out of 1,756,855 arriving flights in 2022.
- The two main contributors to the delays were the airline, which accounted for 42.24% of delays, and security, which accounted for 34.31% of delays.
- Delays due to weather were a relatively small contributor, accounting for only 3.45% of all delays.

NAS=National Airspace System

Arrival Flight Delays Per Airport

(2018-2022)

Number of Arriving Flights per Region (2022)

348,206

207,983

159,511

477,567

WEST

MIDWEST

NORTHEAST

SOUTH

Arriving flights are calculated as the sum of all arriving flights from any airline at the 18 identified airports.
Number in luggage represents number of arriving flights in 2022.
South has 3x Northeast flights.   West has 2.2x Northeast flights.
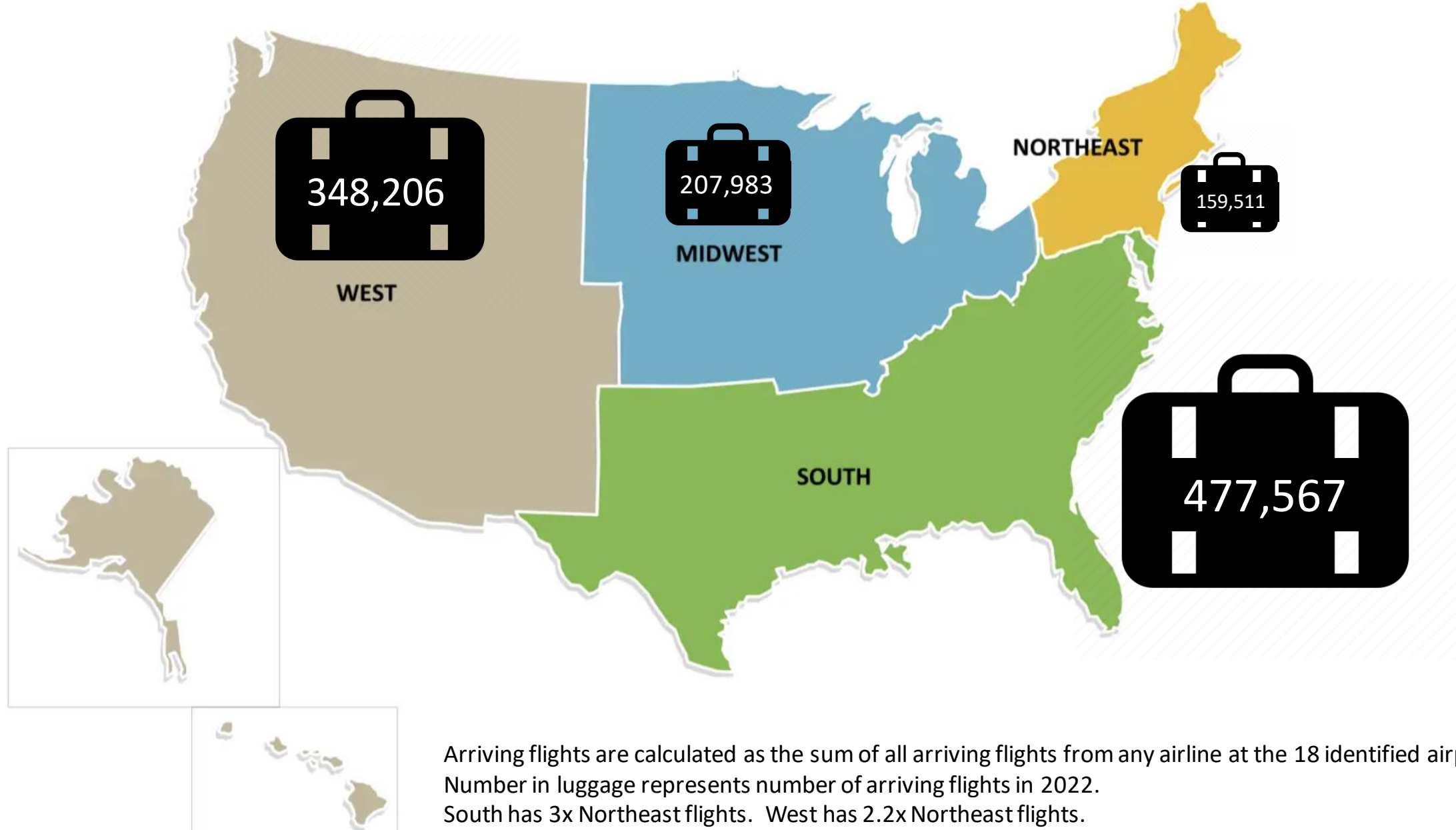
Number of Flights Delayed per Airport (2022)

	Airport	Arrived	Delayed	# Delay Airline	# Delay Weather	% Any Delayed	% Delayed d/t Airline	% Delayed d/t Weather
North	EWR	54130	15198	3671.61	376.46	28.08	24.16	2.48
	LGA	70206	15922	4233.57	568.1	22.68	26.59	3.57
	PHL	35175	7278	2659.88	201.11	20.69	36.55	2.76
	Total	**159511**	**38398**			**23.82**		**2.94**
South	ATL	129219	19885	7248.98	629.43	15.39	36.45	3.17
	BWI	33654	7139	2395.81	120	21.21	33.56	1.68
	CLT	78425	12402	3925.92	561.15	15.81	31.66	4.52
	DFW	112280	18602	6126.75	1040.05	16.57	32.94	5.59
	IAD	22797	4185	1716.11	167.87	18.36	41.01	4.01
	IAH	52568	9048	3687.39	354.3	17.21	40.75	3.92
	MIA	48624	11719	3676.48	339.38	24.10	31.37	2.90
	Total	**477567**	**82980**			**18.38**		**3.68**
Midwest	DTW	53273	9592	4418.91	723.93	18.01	46.07	7.55
	MSP	49295	8885	4324.96	871.09	18.02	48.68	9.80
	ORD	105415	19046	6958.7	714.16	18.07	36.54	3.75
	Total	**207983**	**37523**			**18.03**		**7.03**
West	DEN	105809	21144	8276.03	534.04	19.98	39.14	2.53
	LAX	78450	13753	6077.96	507.97	17.53	44.19	3.69
	SEA	67417	11684	4114.05	437.27	17.33	35.21	3.74
	SFO	51670	8149	3548.53	281.88	15.77	43.55	3.46
	SLC	44860	6308	3259.78	378.6	14.06	51.68	6.00
	Total	**348206**	**61038**			**16.93**		**3.88**

The 18 airports listed represent 44% of flights in US in 2022.
Most flight delays in 2022: Denver 21,144 followed by Atlanta (19,885), then Chicago O'Hare (19,046).

Number of Delayed Flights per Airport and Region (2022)

- Delays are 14 to 28% per airport.

- Most flight delays in 2022: Denver 21,144 followed by Atlanta (19,885), then Chicago O'Hare (19,046).

- Newark (28%), Laguardia (22%) and Miami (24%) have highest % Delayed flights
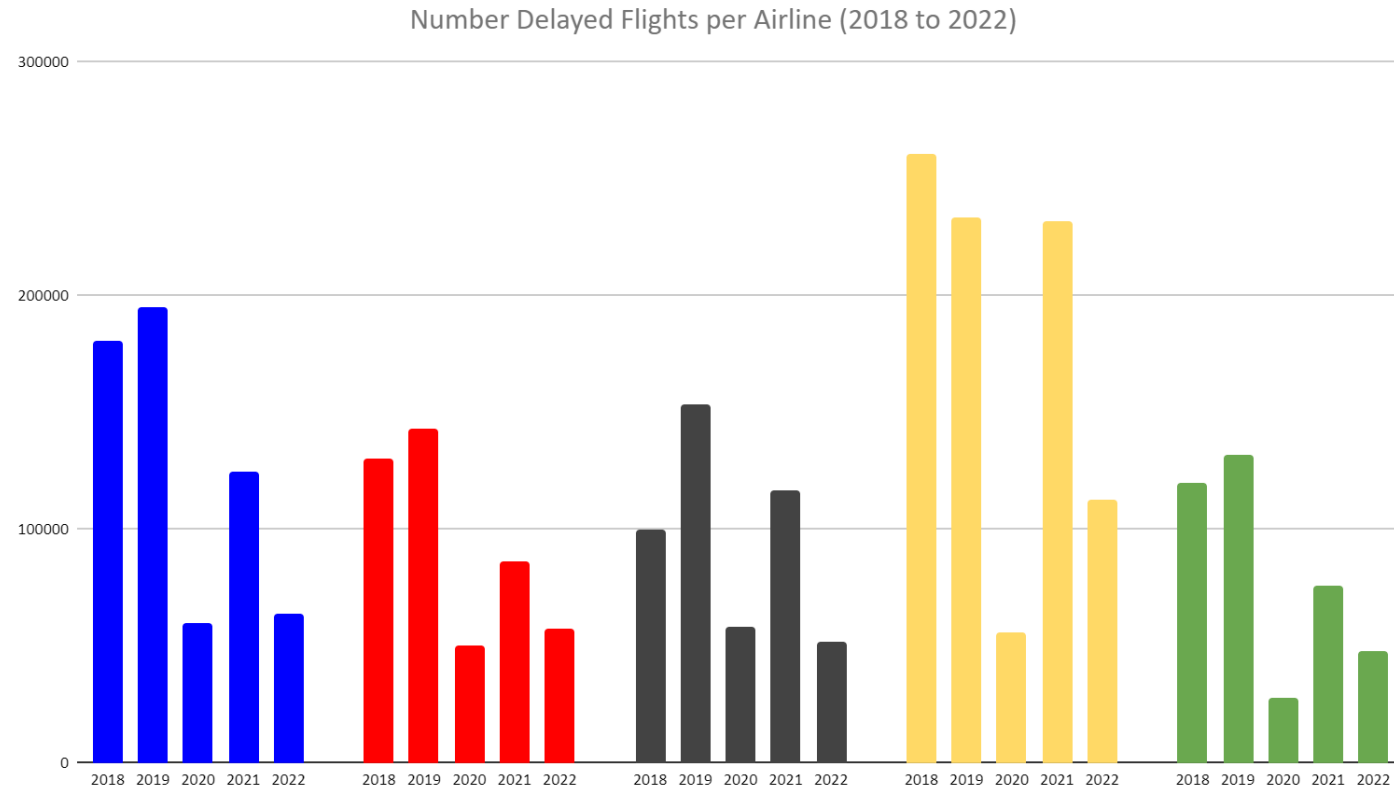
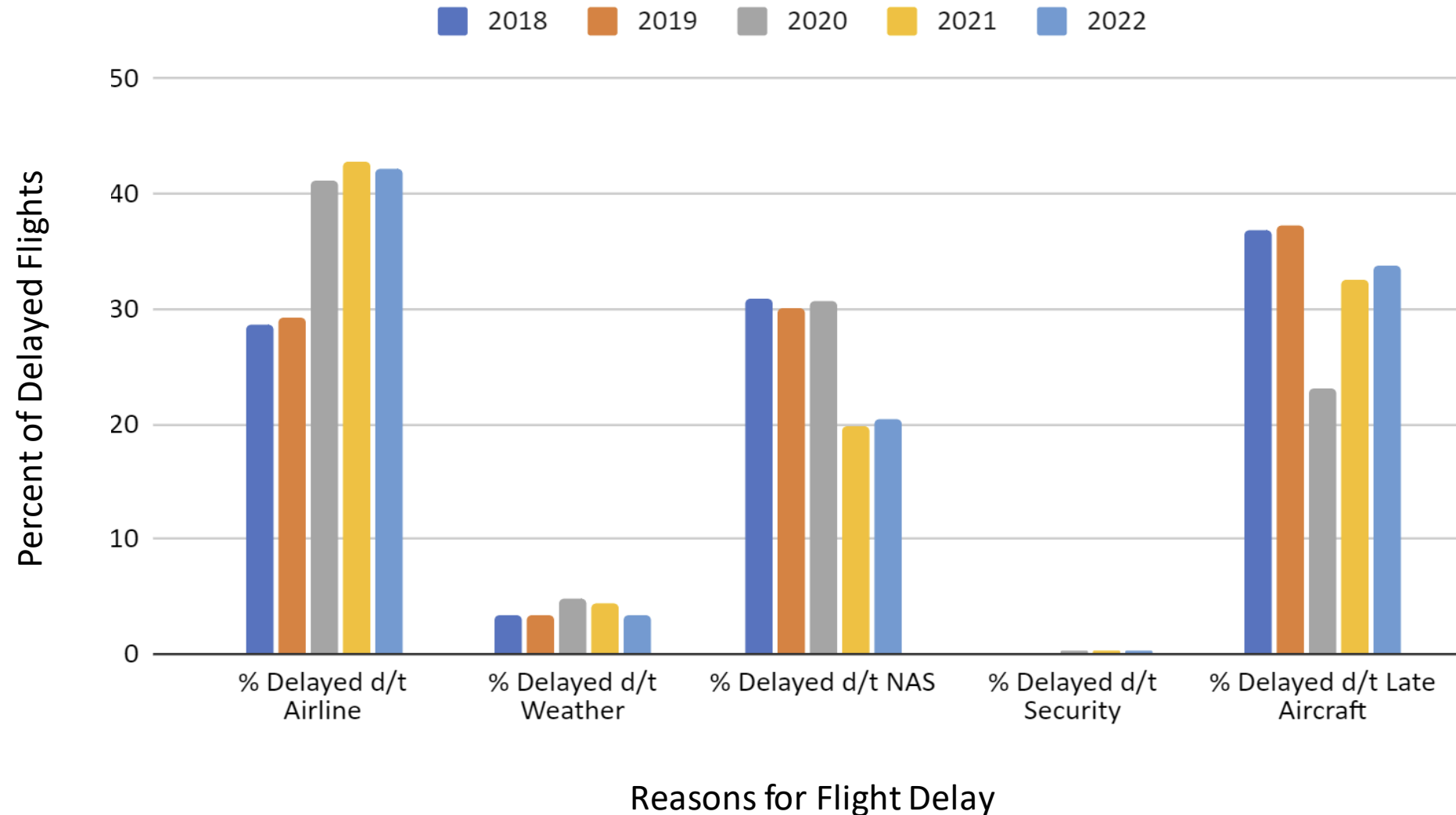# Arrival Flight Delays Per Airline

(2018-2022)

Number of Flights Delayed for Any Reason per Airline (2018-2022)

Year	Arrived	Delayed	# Delay Airline	# Delay Weather	# Delay NAS	# Delay Security	# Delay Late_Plane	% Any Delayed	% Delayed d/t Airline	% Delayed d/t Weather	% Delayed d/t NAS	% Delayed d/t Security	% Delayed d/t Late Aircraft
2018	4371808.0	790548.0	226006.97	27107.08	244985.35	1418.44	291030.28	18.08	28.59	3.43	30.99	0.18	36.81
2019	4765061.0	856478.0	250742.64	28094.33	257784.97	1596.94	318258.88	17.97	29.28	3.28	30.10	0.19	37.16
2020	3017259.0	251010.0	103124.45	12153.75	77105.35	779.56	57846.47	8.32	41.08	4.84	30.72	0.31	23.05
2021	3749286.0	635614.0	272410.20	27734.96	126010.44	2296.64	207162.07	16.95	42.86	4.36	19.82	0.36	32.59
2022	1756855.0	333916.0	141048.63	11527.60	67939.87	857.41	112542.43	19.01	42.24	3.45	20.35	0.26	33.70

Number of Flights Delayed for Any Reason Per Airline (2018-2022)

Number Delayed Flights per Airline (2018 to 2022)

Percent of Delayed Flights For Any Reason as Percent of Total Flight Delays Aggregated for 5 Biggest Airlines (2018-2022)
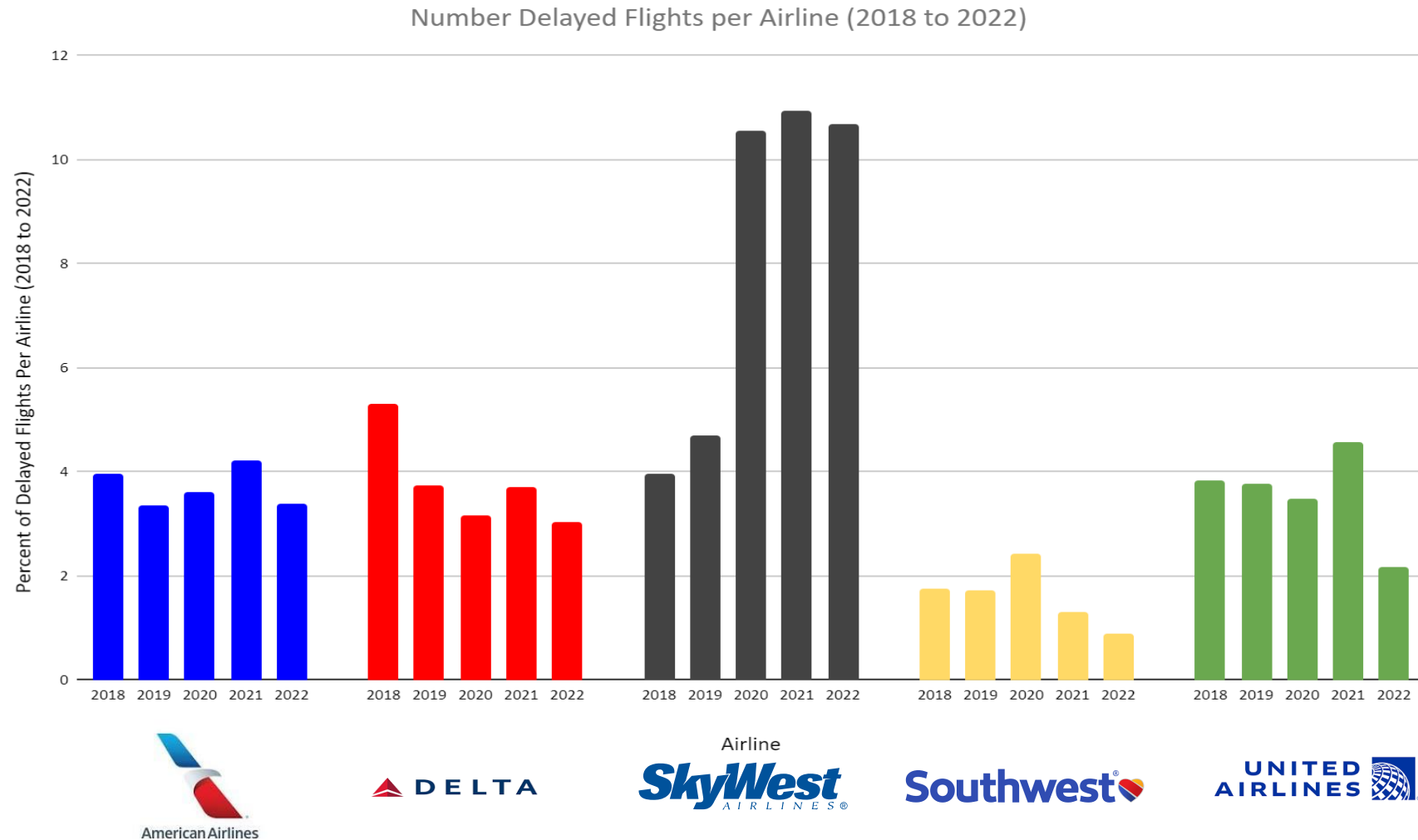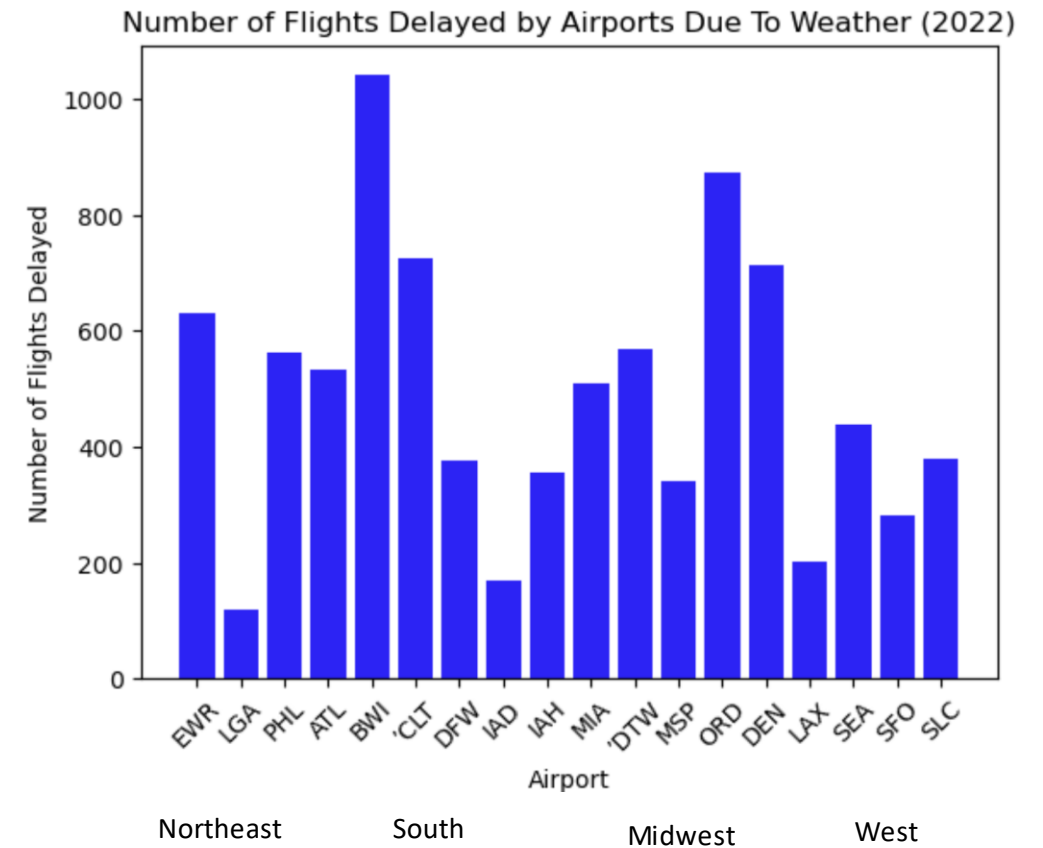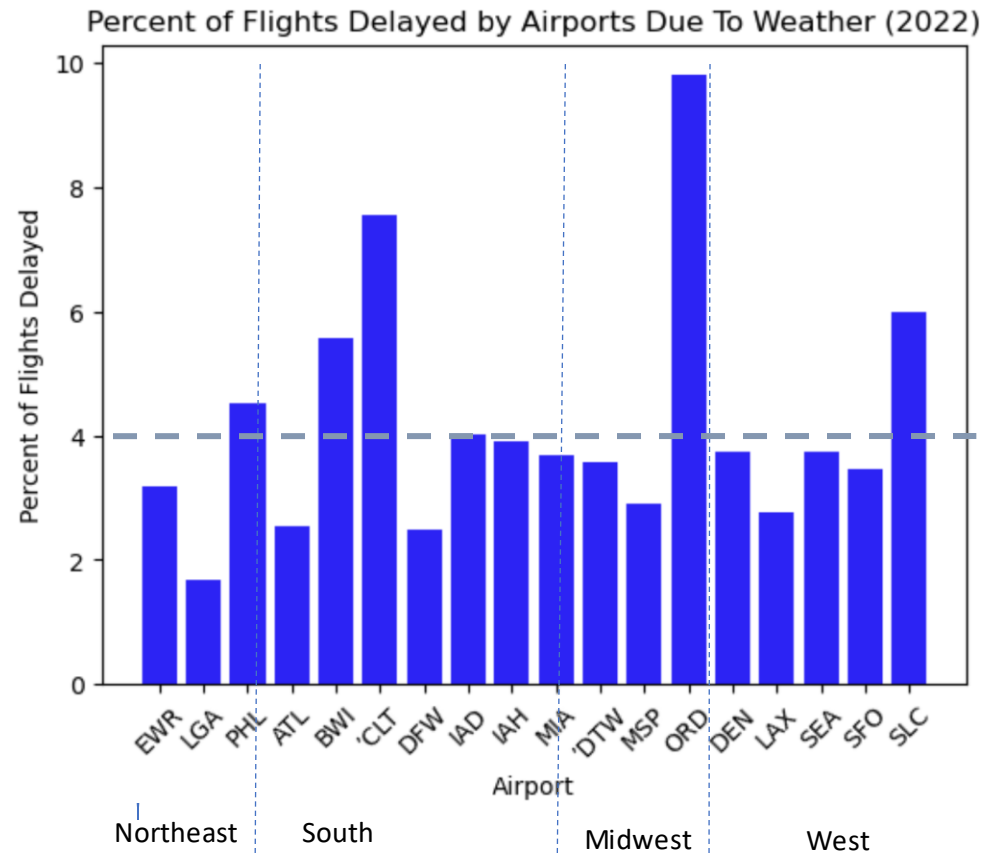
# Impact of Extreme Weather for Flight Delays

(2018-2022)

# Percent of Flights Delayed due to Extreme Weather per Airline (2018-2022)

Number Delayed Flights per Airline (2018 to 2022)

Number (%) of Flights Delayed due to Extreme Weather per Airport (2022)

Delays Due to Weather 2018 to 2022

Percent Delayed Due to Weather at 18 Airports (2018 to 2022)

Delays Due to Weather 2018 to 2020 - West Region
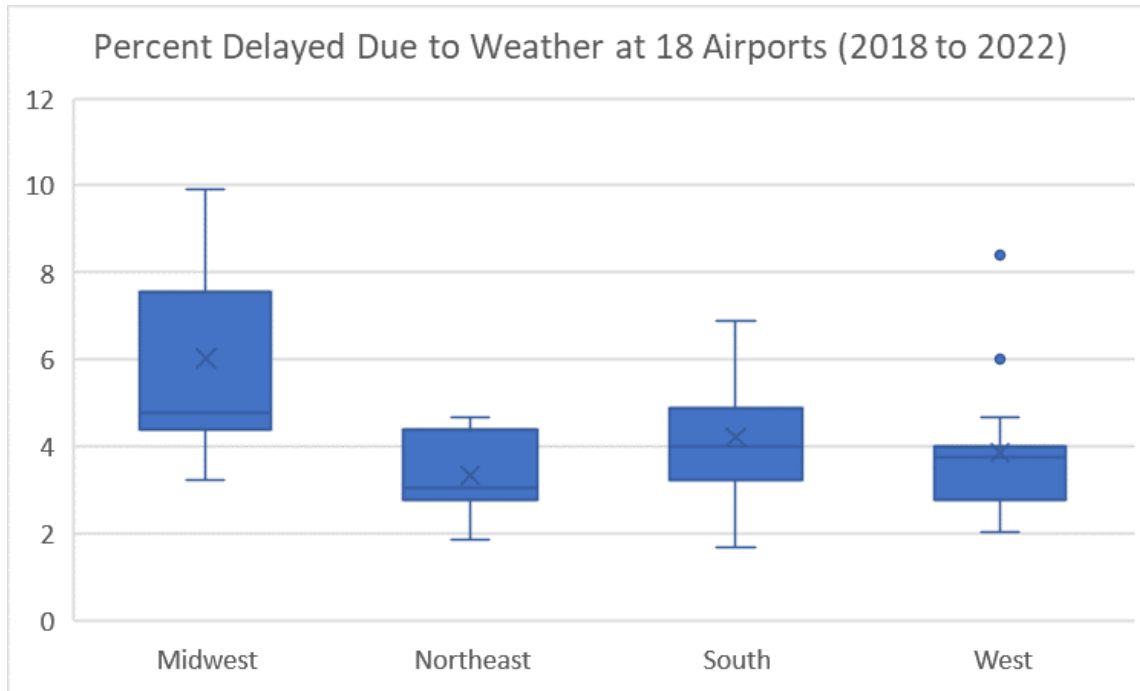
- 3 outliers – Salt Lake City (2020, 2021, 2022)
- Data in Northeast and West are more consistent.
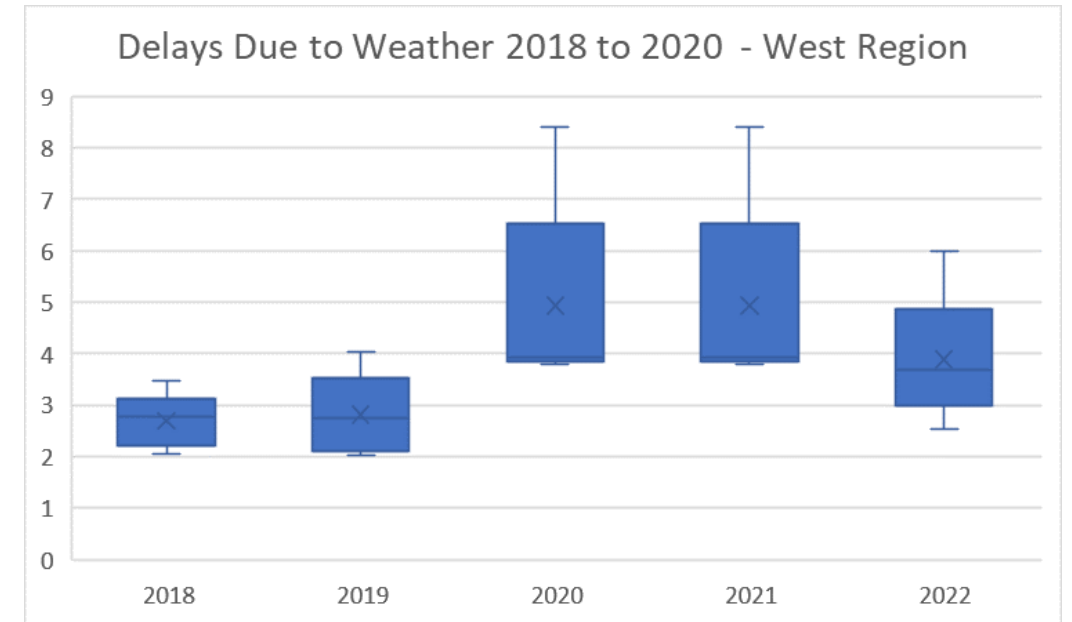
- In West, percent of delays due to weather was higher in 2020, 2021, and 2022 compared to prior years
- Lower rate of delays and more consistent data in 2022 compared to 2020 and 2021, overlapping with rate seen in 2018 and 2019

Airport Hubs For 5 Biggest Airlines

American Airlines	Delta	SouthWest	United	SkyWest
Atlanta - ATL	Atlanta - ATL	Denver - DEN	Chicago - ORD	Chicago - ORD
Dallas/Forth Worth - DFW	Minneapolis – MSP	Baltimore - BWI	Denver - DEN	Denver - DEN
Charlotte - CTL	Detroit - DTW	Chicago - ORD	San Francisco - SFO	Dallas/Fort Worth - DFW
Chicago O'Hare	Salt Lake City - SLC	Atlanta - ATL	Houston - IAH	Salt Lake City - SLC
Miami – MIA		Las Vegas - LAS	Los Angeles - LAX	Los Angeles - LAX
Philadelphia, PHL		Orlando – MCO	Washington DC - IAD	San Francisco - SFO
Phoenix – PHX				

- Most or all the hubs for the 5 biggest airlines experienced weather delays.
- The hubs are located across all 4 regions of the US.

Months With at Least 50 Extreme Weather Related Delays by at least 1 Airline (2018-2022)

|		Region	Total	American Airlines	Delta	Southwest	United	SkyWest
New York	LGA	Northeast	1					1
Newark	EWR	Northeast	3				3	
Philadelphia	PHL	Northeast	4	4				
Atlanta	ATL	South	36		34			2
Baltimore	BWI	South	2			2		
Charlotte	CTL	South	12	12				
Dallas/Forth Worth	DFW	South	39	35				4
Houston	IAH	South	8				8	
Miami	MIA	South	1	1				
Washington DC	IAD	South	1				1	
Chicago	ORD	Midwest	30	9			12	9
Detroit	DTW	Midwest	17				17	
Minneapolis	MSP	Midwest	25		1			24
Denver	DEN	West	28			3	7	18
Loas Angeles	LAX	West	13	4			1	8
Salt Lake City	SLC	West	17					17
San Francisco	SFO	West	5				3	2
Seattle	SEA	West	1					1
				65	**35**	**5**	**35**	**103**

Dallas most frequently experiences weather delays – 39 months of possible 60 months 2018 to 2020

Reduction of Weather Delays Based on US Region (at 18 Airports) (2018 to 2022)
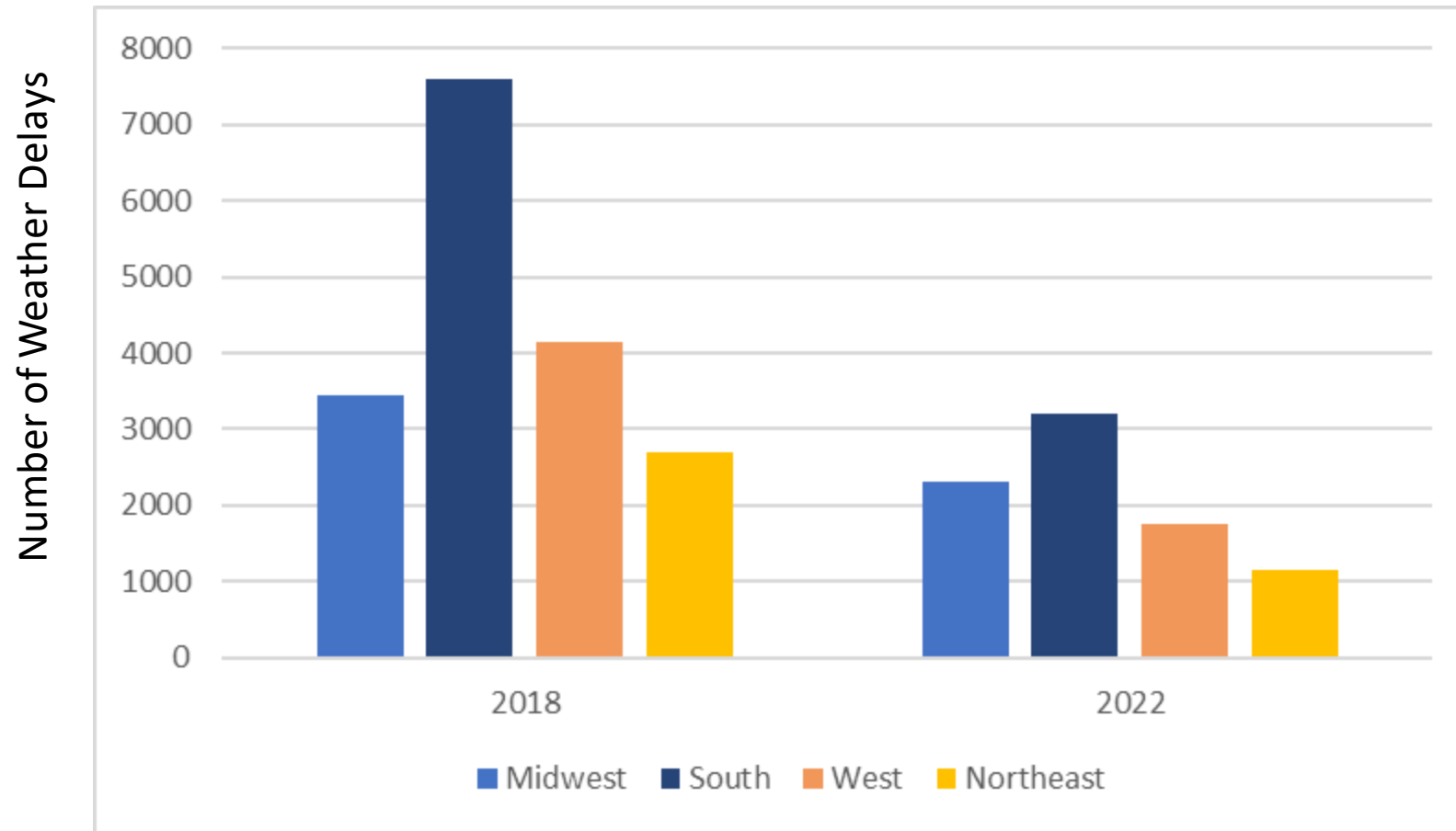
INSIGHTS

- Recent weather changes have caused an **increase in the number of delays** of arrival flights in **all regions** across the US.
  - Delays for all reasons are same or lower vs prior years
  - Weather delays in aggregate have decreased in all US regions (33-58%)
- The **West** and **East Coast** regions have experienced the highest increase in delayed flights.
  - Based on % delays due to weather, delays have decreased back to the rate seen in 2018 (~3.45%)
  - Total number of delayed flights due to weather is less than in 2018 (11.5K vs 27.1K)
- Of the top 5 airlines in the US, **American Airlines** experienced the most delays, cancellations and diversions.
  - Southwest has highest number of delays for any reason
  - SkyWest has the highest number of delays due to weather
  - Delta most frequently has weather delays (most months)
  - American Airlines has weather delays in the highest number of airports

Learnings

- Original data (health-related emergency airline delays) was not readily available after much searching
- Order of code is important
- If data output is not logical, it is good to crosscheck with another tool (eg, Excel) to help refine code in notebook
- Label columns initially at top of the notebook to minimize rework after rest of code were
- Plans to incorporate geo map of delays per region and changed it over time, but we did not have time because we did not merge the dataframe of region with airport code at the outset of the project
- Communication among team is important
- Would have been better to start with a smaller scope (eg, 1 year or 1 region only)