Delay Detectives

Do Different Airlines Have More Frequent Delays?

Table of Contents

I. Introductions

Have you ever been delayed at an airport before? How long was the delay? Did it mess up other plans because of the timing, or did you miss a connecting flight after the delay? Our project hopes to test which airlines are the best at reliably providing accurate flight times. We will be looking at the top five airlines in the United States: American Airlines, Southwest, United, Delta, and Alaska Airlines. These airlines were not randomly selected, so our results can only be generalized to this population of the top five American airlines.

II. Statistical Question

The question we are trying to answer is: Do different airlines have more frequent delays?

Hypotheses:

H_0 : For each of the 5 airlines, the distribution of planes that are “Early-15 mins late”, “16-60 mins late”, and “Over 60 mins late” is the same.

H_A: For each of the 5 airlines, the distribution of planes that are “Early-15 mins late”, “16-60 mins late”, and “Over 60 mins late” is not the same.

III. Data Collection

In order to collect data, we analyzed each of our selected airlines at five randomly selected airports, taking a cluster sample of every departure that airline made over ten randomly selected days of the year. To do this, we first needed to select which airports to analyze. We found data for the 94 largest airports in the US (those with over 1 million enplanements per year), assigning each airport a number from 2-95. We then used a random number generator 2-95 to generate five unique numbers from that interval. Those five values were 82, 3, 24, 81, 57. From our spreadsheet of airport data, these corresponded to the following airports: Albany

International Airport, Los Angeles International Airport, Salt Lake City International Airport, Tulsa International Airport, and Hollywood Burbank (Bob Hope) Airport.

We then needed to randomly select the days that we are collecting data from. We used reports from the year 2021 because that was the most recent full year with available data. This gives us the most accurate and up-to-date data. Similar to the airlines, we used a random number generator from 1-365 to generate ten unique numbers, for ten unique days of the year. Those random numbers were: 66, 271, 172, 248, 171, 326, 166, 345, 324, 118. We then choose to collect data from days corresponding to these numbers: March 7, September 28, June 21, September 5, June 20, November 22, June 15, December 11, November 20, and April 28.

State	City	FAA	IATA	ICAO	Airport	Role	Enplanements
GEORGIA	Atlanta	ATL	ATL	KATL	Hartsfield–Jackson Atlanta International Airport	P-L	51,865,797
CALIFORNIA	Los Angeles	LAX	LAX	KLAX	Los Angeles International Airport	P-L	42,624,050
ILLINOIS	Chicago	ORD	ORD	KORD	Chicago O'Hare International Airport	P-L	39,873,927
TEXAS	Dallas	DFW	DFW	KDFW	Dallas Fort Worth International Airport	P-L	32,821,799
COLORADO	Denver	DEN	DEN	KDEN	Denver International Airport	P-L	31,362,941
NEW YORK	New York	JFK	JFK	KJFK	John F. Kennedy International Airport (was New York International Airport)	P-L	30,620,769
CALIFORNIA	San Francisco	SFO	SFO	KSFO	San Francisco International Airport	P-L	27,790,717
WASHINGTON	Seattle / Tacoma (SeaTac)	SEA	SEA	KSEA	Seattle–Tacoma International Airport	P-L	24,024,908
NEVADA	Las Vegas	LAS	LAS	KLAS	Harry Reid International Airport	P-L	23,795,012
FLORIDA	Orlando	MCO	MCO	KMCO	Orlando International Airport	P-L	23,202,480
NEW JERSEY	Newark	EWR	EWR	KEWR	Newark Liberty International Airport	P-L	22,797,602
NORTH CAROLINA	Charlotte	CLT	CLT	KCLT	Charlotte Douglas International Airport	P-L	22,281,949
ARIZONA	Phoenix	PHX	PHX	KPHX	Phoenix Sky Harbor International Airport	P-L	21,662,580
TEXAS	Houston	IAH	IAH	KIAH	George Bush Intercontinental/Houston Airport	P-L	21,157,398
FLORIDA	Miami	MIA	MIA	KMIA	Miami International Airport	P-L	21,021,640
MASSACHUSETTS	Boston	BOS	BOS	KBOS	Gen. Edward Lawrence Logan International Airport	P-L	20,006,521
MINNESOTA	Minneapolis-St. Paul	MSP	MSP	KMSP	Minneapolis–St. Paul International/Wold-Chamberlain Airport	P-L	18,361,942
FLORIDA	Fort Lauderdale	FLL	FLL	KFLL	Fort Lauderdale–Hollywood International Airport	P-L	17,612,331
MICHIGAN	Detroit	DTW	DTW	KDTW	Detroit Metro Wayne County Airport	P-L	17,436,837
PENNSYLVANIA	Philadelphia	PHL	PHL	KPHL	Philadelphia International Airport	P-L	15,292,670
NEW YORK	New York	LGA	LGA	KLGA	LaGuardia Airport (and Marine Air Terminal)	P-L	15,058,501
MARYLAND	Baltimore	BWI	BWI	KBWI	Baltimore/Washington International Airport	P-L	13,371,816
UTAH	Salt Lake City	SLC	SLC	KSLC	Salt Lake City International Airport	P-L	12,226,730
CALIFORNIA	San Diego	SAN	SAN	KSAN	San Diego International Airport (Lindbergh Field)	P-L	12,174,224
VIRGINIA	Washington, D.C. / Dulles / Chantilly	IAD	IAD	KIAD	Washington Dulles International Airport	P-L	11,621,623
VIRGINIA	Washington, D.C. / Arlington	DCA	DCA	KDCA	Ronald Reagan Washington National Airport	P-L	11,366,771
ILLINOIS	Chicago	MDW	MDW	KMDW	Chicago Midway International Airport	P-L	10,678,018
FLORIDA	Tampa	TPA	TPA	KTPA	Tampa International Airport	P-L	10,378,514
HAWAII	Honolulu, Oahu	HNL	HNL	PHNL	Daniel K. Inouye International Airport	P-L	10,017,149
OREGON	Portland	PDX	PDX	KPDX	Portland International Airport	P-L	9,804,868
TENNESSEE	Nashville	BNA	BNA	KBNA	Nashville International Airport (Berry Field)	P-M	8,017,347
TEXAS	Dallas	DAL	DAL	KDAL	Dallas Love Field	P-M	8,011,221
TEXAS	Austin	AUS	AUS	KAUS	Austin–Bergstrom International Airport	P-M	7,714,479
MISSOURI	St. Louis	STL	STL	KSTL	St. Louis Lambert International Airport	P-M	7,631,953
TEXAS	Houston	HOU	HOU	KHOU	William P. Hobby Airport	P-M	7,053,886
CALIFORNIA	San Jose	SJC	SJC	KSJC	Norman Y. Mineta San José International Airport	P-M	7,032,851
CALIFORNIA	Oakland	OAK	OAK	KOAK	Oakland International Airport	P-M	6,686,603
LOUISIANA	New Orleans	MSY	MSY	KMSY	Louis Armstrong New Orleans International Airport	P-M	6,582,066
NORTH CAROLINA	Raleigh	RDU	RDU	KRDU	Raleigh–Durham International Airport	P-M	6,258,101
CALIFORNIA	Sacramento	SMF	SMF	KSMF	Sacramento International Airport	P-M	5,907,629
MISSOURI	Kansas City	MCI	MCI	KMCI	Kansas City International Airport (was Mid-Continent International)	P-M	5,790,847
CALIFORNIA	Orange County	SNA	SNA	KSNA	John Wayne Airport (was Orange County Airport)	P-M	5,201,642
TEXAS	San Antonio	SAT	SAT	KSAT	San Antonio International Airport	P-M	4,844,427
OHIO	Cleveland	CLE	CLE	KCLE	Cleveland Hopkins International Airport	P-M	4,701,713
PENNSYLVANIA	Pittsburgh	PIT	PIT	KPIT	Pittsburgh International Airport	P-M	4,670,033
INDIANA	Indianapolis	IND	IND	KIND	Indianapolis International Airport	P-M	4,655,847
FLORIDA	Fort Myers	RSW	RSW	KRSW	Southwest Florida International Airport	P-M	4,602,113
KENTUCKY	Cincinnati/Covington	CVG	CVG	KCVG	Cincinnati/Northern Kentucky International Airport	P-M	4,269,258
PUERTO RICO	San Juan / Carolina	SJU	SJU	TJSJ	Luis Muñoz Marín International Airport	P-M	4,033,412
OHIO	Columbus	CMH	CMH	KCMH	John Glenn Columbus International Airport	P-M	3,976,620
HAWAII	Kahului, Maui	OGG	OGG	PHOG	Kahului Airport	P-M	3,571,660
WISCONSIN	Milwaukee	MKE	MKE	KMKE	Milwaukee Mitchell International Airport	P-M	3,496,724
CONNECTICUT	Hartford	BDL	BDL	KBDL	Bradley International Airport	P-M	3,378,035
FLORIDA	West Palm Beach	PBI	PBI	KPBI	Palm Beach International Airport	P-M	3,270,605
FLORIDA	Jacksonville	JAX	JAX	KJAX	Jacksonville International Airport	P-M	3,135,508
CALIFORNIA	Burbank	BUR	BUR	KBUR	Hollywood Burbank Airport (was Bob Hope Airport)	P-M	2,680,240
NEW MEXICO	Albuquerque	ABQ	ABQ	KABQ	Albuquerque International Sunport	P-M	2,647,269
ALASKA	Anchorage	ANC	ANC	PANC	Ted Stevens Anchorage International Airport	P-M	2,642,607
NEW YORK	Buffalo	BUF	BUF	KBUF	Buffalo Niagara International Airport	P-M	2,523,158
CALIFORNIA	Ontario	ONT	ONT	KONT	Ontario International Airport	P-M	2,498,993
NEBRASKA	Omaha	OMA	OMA	KOMA	Eppley Airfield	P-M	2,454,878
TENNESSEE	Memphis	MEM	MEM	KMEM	Memphis International Airport	P-S	2,213,083
SOUTH CAROLINA	Charleston	CHS	CHS	KCHS	Charleston International Airport / Charleston AFB	P-S	2,192,893
RHODE ISLAND	Providence	PVD	PVD	KPVD	Rhode Island T. F. Green International Airport	P-S	2,117,409
OKLAHOMA	Oklahoma City	OKC	OKC	KOKC	Will Rogers World Airport	P-S	2,094,708
NEVADA	Reno	RNO	RNO	KRNO	Reno/Tahoe International Airport	P-S	2,048,916
VIRGINIA	Richmond	RIC	RIC	KRIC	Richmond International Airport (Byrd Field)	P-S	2,048,691
IDAHO	Boise	BOI	BOI	KBOI	Boise Airport (Boise Air Terminal) (Gowen Field)	P-S	1,943,181
CALIFORNIA	Long Beach	LGB	LGB	KLGB	Long Beach Airport (Daugherty Field)	P-S	1,908,635
KENTUCKY	Louisville	SDF	SDF	KSDF	Louisville International Airport (Standiford Field)	P-S	1,877,861
WASHINGTON	Spokane	GEG	GEG	KGEG	Spokane International Airport (Geiger Field)	P-S	1,872,781
VIRGINIA	Norfolk	ORF	ORF	KORF	Norfolk International Airport	P-S	1,846,031
HAWAII	Kailua-Kona, Hawaii	KOA	KOA	PHKO	Ellison Onizuka Kona International Airport at Keahole	P-S	1,829,020
ARIZONA	Tucson	TUS	TUS	KTUS	Tucson International Airport	P-S	1,753,227
GUAM	Agana / Tamuning	GUM	GUM	PGUM	Antonio B. Won Pat International Airport	P-S	1,736,131
HAWAII	Lihue, Kauai	LIH	LIH	PHLI	Lihue Airport	P-S	1,644,590
MICHIGAN	Grand Rapids	GRR	GRR	KGRR	Gerald R. Ford International Airport	P-S	1,631,398
TEXAS	El Paso	ELP	ELP	KELP	El Paso International Airport	P-S	1,605,486
FLORIDA	Sanford	SFB	SFB	KSFB	Orlando Sanford International Airport	P-S	1,504,888
OKLAHOMA	Tulsa	TUL	TUL	KTUL	Tulsa International Airport	P-S	1,482,908
NEW YORK	Albany	ALB	ALB	KALB	Albany International Airport	P-S	1,440,674
GEORGIA	Savannah	SAV	SAV	KSAV	Savannah/Hilton Head International Airport	P-S	1,356,660
IOWA	Des Moines	DSM	DSM	KDSM	Des Moines International Airport	P-S	1,347,876
NEW YORK	Rochester	ROC	ROC	KROC	Frederick Douglass/Greater Rochester International Airport	P-S	1,281,908
SOUTH CAROLINA	Myrtle Beach	MYR	MYR	KMYR	Myrtle Beach International Airport	P-S	1,254,307
CALIFORNIA	Palm Springs	PSP	PSP	KPSP	Palm Springs International Airport	P-S	1,163,883
NEW YORK	Syracuse	SYR	SYR	KSYR	Syracuse Hancock International Airport	P-S	1,139,568
SOUTH CAROLINA	Greenville	GSP	GSP	KGSP	Greenville–Spartanburg International Airport (Roger Milliken Field)	P-S	1,133,012
FLORIDA	St. Petersburg	PIE	PIE	KPIE	St. Pete–Clearwater International Airport	P-S	1,115,886
ALABAMA	Birmingham	BHM	BHM	KBHM	Birmingham–Shuttlesworth International Airport	P-S	1,081,328
TENNESSEE	Knoxville	TYS	TYS	KTYS	McGhee Tyson Airport	P-S	1,069,565
MAINE	Portland	PWM	PWM	KPWM	Portland International Jetport	P-S	1,062,873
WISCONSIN	Madison	MSN	MSN	KMSN	Dane County Regional Airport (Truax Field)	P-S	1,043,185
ARKANSAS	Little Rock	LIT	LIT	KLIT	Bill and Hillary Clinton National Airport (Adams Field) (was Little Rock National)	P-S	1,031,258

Now that we have the sample airports and the sample dates, we must create our delay categories. The Federal Aviation Administration (FAA) considers a flight to be delayed when it is over 15 minutes later than its scheduled departure time. However, our group also wanted to

include a category for delays exceeding an hour. So, we will check each airline to see how many flights truly departed on time, how many were delayed 16-60 minutes, and how many were delayed over an hour. We can generalize our findings about these specific airlines to every airport with at least 1,000,000 enplanements in the US, because those were the airports that were randomly sampled from. Our results will determine which airlines have the most consistent departure times, and which airlines you likely shouldn't rely on for timely flights.

We collected the data from the Bureau of Transportation Statistics. The act of collecting data was a long and tedious process, the website was only allowed to search for one airline, and one day at a time, meaning we needed two hundred and fifty different searches with every statistic needing to be manually placed into a spreadsheet. At the end of this process we had successfully collected 5507 flights from those airlines, airports, and days from 2021.

IV. Data Displays

*The following data is collected from one day (April 28th). It is one of ten days we collected data on.

Flights that were early - 15 minutes late:

# of Planes	American	Delta	United	Southwest	Alaska	April 28
Albany International	2	3	1	6	0	
Los Angeles International	51	62	28	34	21	
Salt Lake City International	1	106	5	17	1	
Tulsa International	4	1	3	4	0	
Hollywood Burbank	1	0	0	16	3	

Flights that were 15 - 60 minutes late:

# of Planes	American	Delta	United	Southwest	Alaska	April 28
Albany International	1	0	0	0	0	
Los Angeles International	7	2	1	2	2	
Salt Lake City International	0	7	0	2	0	
Tulsa International	0	0	0	1	0	
Hollywood Burbank	0	0	0	0	0	

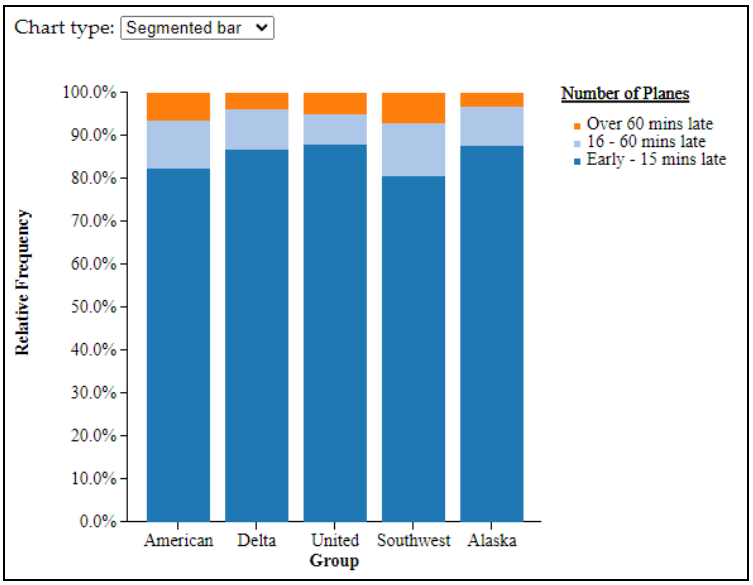
Flights that were > 60 minutes late:

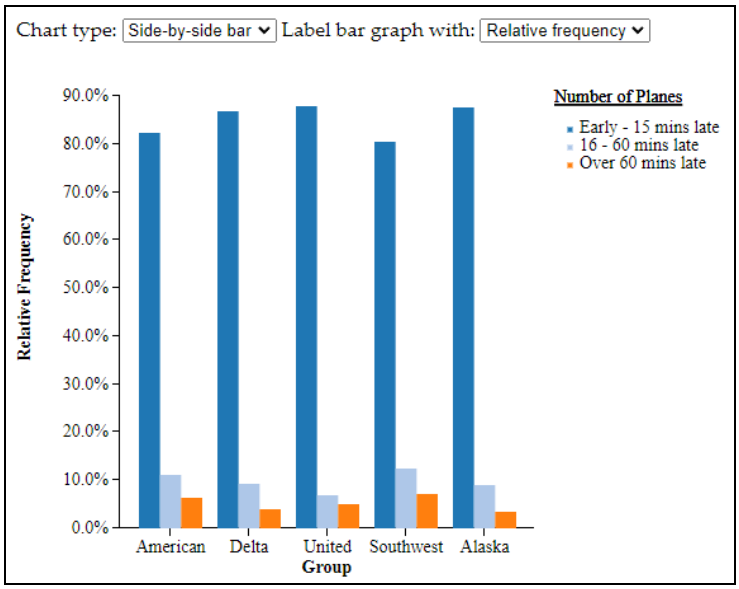
# of Planes	American	Delta	United	Southwest	Alaska	April 28
Albany International	0	0	0	1	0	
Los Angeles International	1	3	1	2	1	
Salt Lake City International	1	2	0	0	0	
Tulsa International	0	0	0	2	0	
Hollywood Burbank	0	0	0	0	0	

Below is the accumulated data for all ten days.

Summary Statistics

		Group					
		American	Delta	United	Southwest	Alaska	**Total**
Number of Planes	Early - 15 mins late	866 (82.4%)	1853 (86.8%)	533 (88%)	1112 (80.5%)	291 (87.7%)	4655 (84.5%)
	16 - 60 mins late	118 (11.2%)	200 (9.4%)	42 (6.9%)	171 (12.4%)	30 (9%)	561 (10.2%)
	Over 60 minutes late	67 (6.4%)	83 (3.9%)	31 (5.1%)	99 (7.2%)	11 (3.3%)	291 (5.3%)
	Total	1051 (100%)	2136 (100%)	606 (100%)	1382 (100%)	332 (100%)	**5507 (100%)**





V. Data Analysis

As you can see in the segmented bar graph and the side-by-side bar graph, the majority of flights in all five of these airlines are either early or considered to be on time (less than fifteen minutes late). Southwest has the highest percentage of very late flights (more than an hour behind scheduled departure time), which occurred 99 times out of the 1382 total Southwest flights (7.16%). The airline with the lowest percentage of very delayed flights is Alaskan Airlines with only 3.3% of all recorded flights being delayed over an hour.

But are these results of the difference in percentages solely due to random chance? Or is there a statistically significant difference in the percentage of flights that are delayed between the five airlines? We will have to perform an inference procedure to make a claim.

VI. Inference Procedure

We will be conducting a Chi-square test for homogeneity in order to determine whether there is a difference in the distribution of planes that are “Early-15 mins late”, “16-60 mins late”,

and “Over 60 mins late” for each airline. We will be using an significance level of $\alpha = 0.05$ to test for significance. But first, we must check the conditions for inference.

Random:

We need the sample to be randomly selected from each population if we want to generalize the results to a larger population of airports and all days of the year. We are randomly sampling five airports from the 94 top airports in the United States and ten days out of the 365 in a year using a random number generator, so the random condition is satisfied.

Independence:

Because we are sampling without replacement from a finite population, we need to ensure that the independence condition is met. The five airports selected are less than 10% of the 94 we are sampling from, and the ten days selected are less than 10% of the 365 possible days in a year that could be selected, so the 10% condition is checked. In terms of individual flights, we conducted a census, and therefore this condition does not apply.

Expected/Large Counts:

The expected counts condition must be met to perform any of the chi-square tests. As shown in the table below, all of the expected counts for airlines and delay intervals are greater than five, so the condition is met.

Expected counts:

	Groups					
Number of Planes		American	Delta	United	Southwest	Alaska
	Early - 15 mins late	888.397	1805.535	512.244	1168.188	280.636
	16 - 60 mins late	107.066	217.595	61.733	140.785	33.821
	Over 60 minutes late	55.537	112.87	32.022	73.027	17.543

All conditions have been met so we are now able to perform the inference procedure. After performing the chi-square test for homogeneity, we observe a test statistic of 43.484, 8 degrees of freedom, resulting in a p-value of <0.001.

Perform Inference

Perform chi-square test for homogeneity

χ^2	P-value	df
43.484	<0.001	8

VII. Conclusions

Because the resulting P-value of <0.001 is less than the significance level of $\alpha = 0.05$, we reject the null hypothesis. There is convincing evidence that for each of the sampled airlines, the distribution of planes that are “Early-15 mins late”, “16-60 mins late”, and “Over 60 mins late” is not the same. From the graphs, contributions to the x^2 value, and summary statistic percentages, we can analyze what airlines have good and bad delay rates. Southwest is most likely to have a delayed flight (16-60 minute delay + over 60 minute delay), followed by American, Delta, Alaska, and United in that order.

Contributions:

	Groups				
	American	Delta	United	Southwest	Alaska
Number of Planes					
Early - 15 mins late	0.565	1.248	0.841	2.703	0.383
16 - 60 mins late	1.117	1.423	6.308	6.485	0.432
Over 60 minutes late	2.366	7.905	0.033	9.237	2.441

VIII. Reflections

Looking back, using 94 airports may have skewed some of the results. We could have been more picky about which airports we sampled from because the number of enplanements that the 2nd largest airport in the US (Los Angeles International Airport, CA) has is much greater

than the number of enplanements that the 82nd largest airport in the US (Albany International Airport, NY) has. Perhaps next time rather than using a list of 94 airports, we could have used a different cutoff value for what was quantified as the "largest airports". Having only 40 airports instead of 94 would reduce variability in the amount of flights from each airport, and make sure each airport is represented with equal weight. Another issue that we ran into was the problem of actually collecting the data, as it was time consuming. In the future, we would do more research to find a more efficient method of collecting data before starting.

Using a stratified random sample for delay times from the top airports would have been more representative of the population than using a cluster sample for all the delay times from a few random airports. But for our data gathering procedures, it was too impractical to collect all the samples with this method. Additionally, collecting a bigger sample from more airports and more days would create more accurate findings, which would increase the power. There could also have been confounding variables with many airports using a majority of one specific airline.

Works Cited

Flightradar24. "Live Flight Tracker - Real-Time Flight Tracker Map." *Flightradar24*, https://www.flightradar24.com/data/airports/lax/departures

Detailed Statistics Departures, https://www.transtats.bts.gov/ontime/departures.aspx

"List of Largest Airlines in North America." *Wikipedia*, Wikimedia Foundation, 21 Nov. 2022, https://en.wikipedia.org/wiki/List_of_largest_airlines_in_North_America

John K, Jack Z, Camden S, Scott B. *Asa Data Collection Airports*, Google, https://docs.google.com/spreadsheets/d/1PxiRqLlWf0Htd0NmlnbgXyCKh4UKwDbNRnbjJQwi5m8/edit#gid=0