Project Process Book

Aditi Ravindra, Yash Kamoji, Sreevidya Bollineni

AIRPORT DELAY ANALYSIS

Links

Github

https://github.com/ykamoji/airport-delay-analysis

Website

https://ykamoji.github.io/airport-delay-analysis/

Initial Project Ideas - Motivation

After browsing through a variety of datasets online, we decided that we would like to work with a dataset with information on flight delays. Our group chose to do this project on Flight Delay Analysis because all three of us frequently take flights as a form of travel, as we are not from Massachusetts. Flight delays are a very frustrating and unfortunately, common experience, which made this topic relevant to us personally and thus, more meaningful. Taking a deeper look into the reason and patterns behind delays with flights felt like an interesting and practical way to apply data analysis to a real-world problem we all regularly face. This project combines both our personal experiences as well as our drive for data-driven exploration; by examining these patterns in delays, we are eager to gain insights that can hopefully help frequent fliers, such as ourselves, better anticipate delays and plan itineraries that go hand in hand with that.

Initial Project Ideas - Objectives

The goal of our project is to analyze flight delay data to identify common causes of delays based on factors such as location, time, and season. By visualizing this data, we want to answer key questions such as the most frequent reasons for delay, the percentage of delayed flights, any time or seasonal patterns that affect the punctuality of flight arrivals, and most dependable airline. Understanding these factors can help not only help passengers anticipate delays, but also allow airlines to be prepared and mitigate them more effectively. Overall, the insights we can gain through analyzing this flight data can help improve flight scheduling within airports, allow for better resource allocation, and enhance customer satisfaction among airlines, all of which contribute to a more efficient travel process.

Related Work

Data Source:

The primary data source and inspiration is the Bureau of Transportation Statistics (BTS), part of the U.S. Department of Transportation, which provides the "Airline On-Time Performance Data". This dataset includes detailed records on departure/arrival times and causes of delay reported by U.S. air carriers.

Data source link: *https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp*

Initial Project Ideas - Visualizations Goals

After combing through a list of ideas that we came up with through our own experiences, we decided that these are the questions that we wanted to answer without our visualizations:

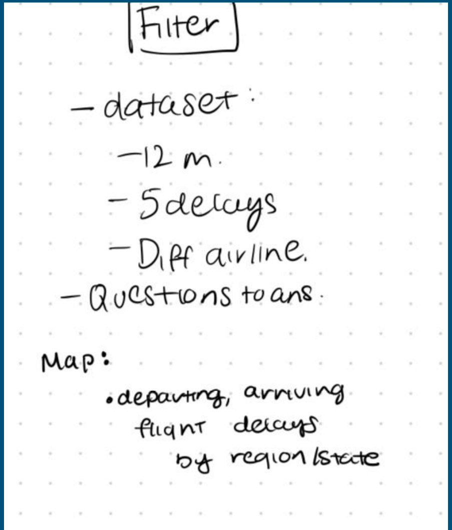
- Which states have the most delays with arriving and departing flights?
- What are the most common causes of delay from the list (5) given?
- What are the most/least efficient airports to travel from?
- What times/months during the year have the most delays? How much are flights delayed by, on average, during each time period throughout the year?
- Are there more delays on weekdays or weekends?

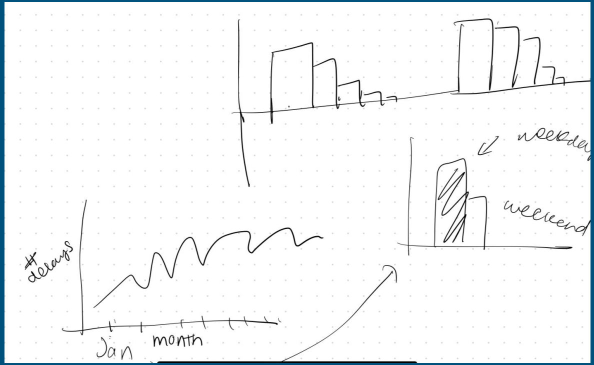
How to Visualize? … Some ideas we had!

We reviewed the dataset and brainstormed several visualization ideas.

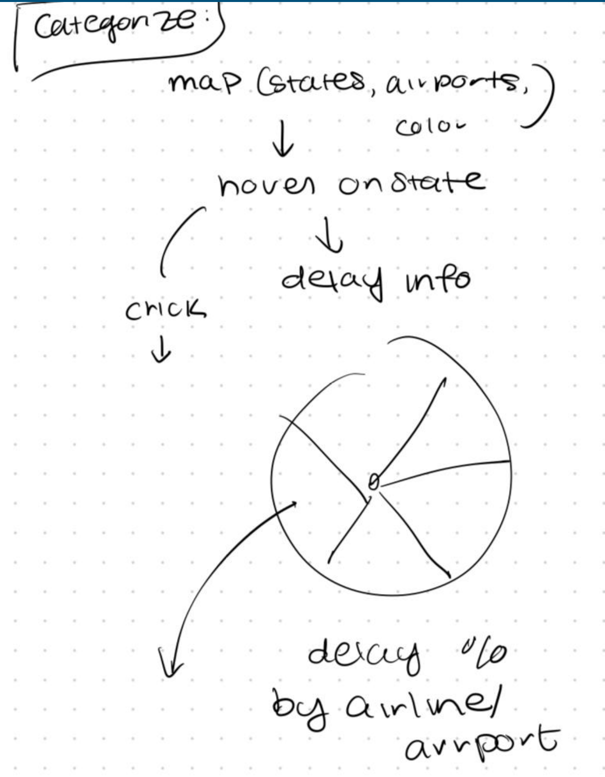
As we explored these options, we identified relevant data subsets tailored for each individual chart.

How to Visualize? … Some ideas we had!



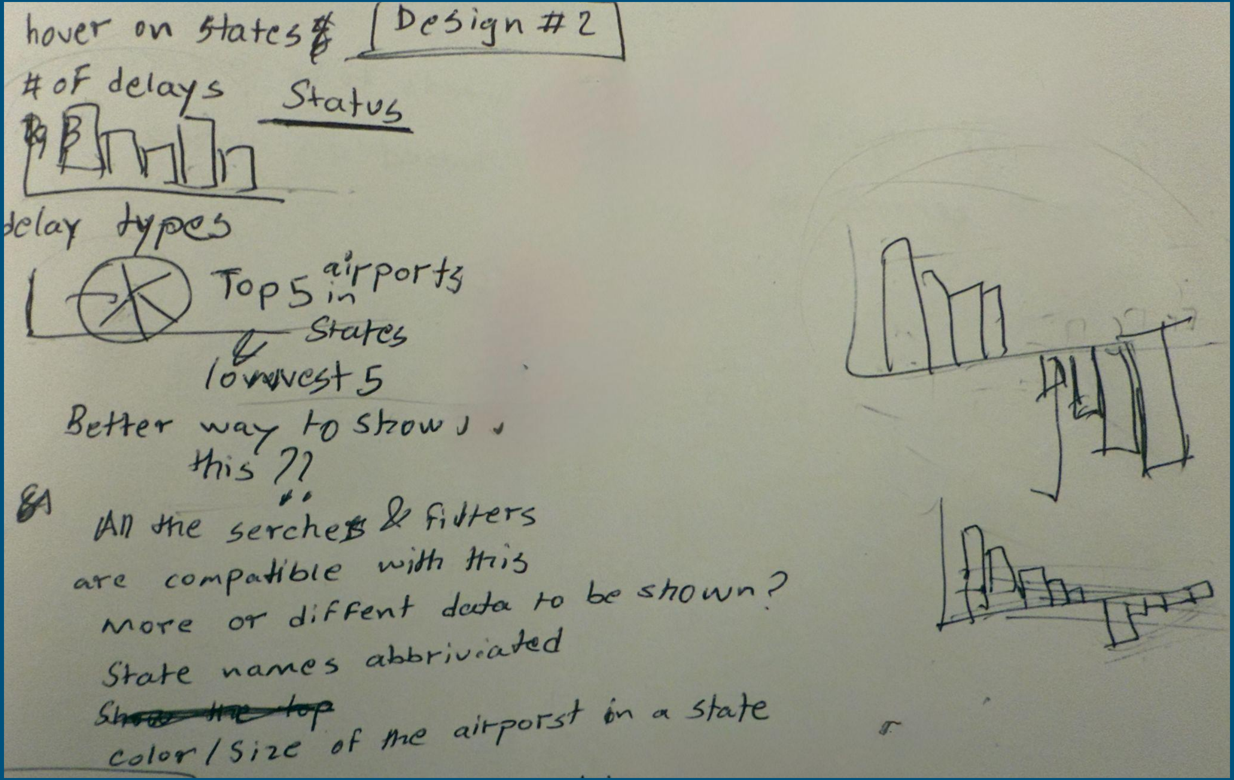


Trying out different ways for the 2nd & 3rd visualization.



We kept few possible options for later.

How to Visualize? … Some ideas we had!



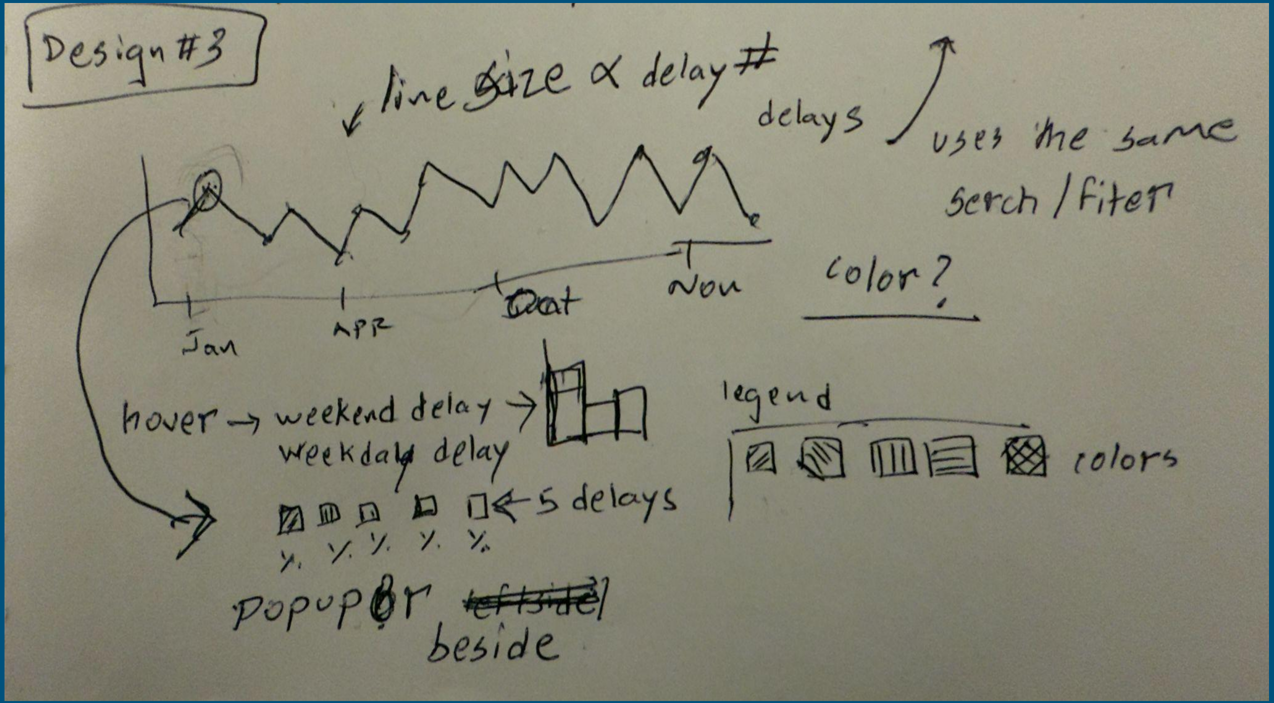
For Design #2, we explored control and interaction options, aiming for a balance between simplicity and functionality.

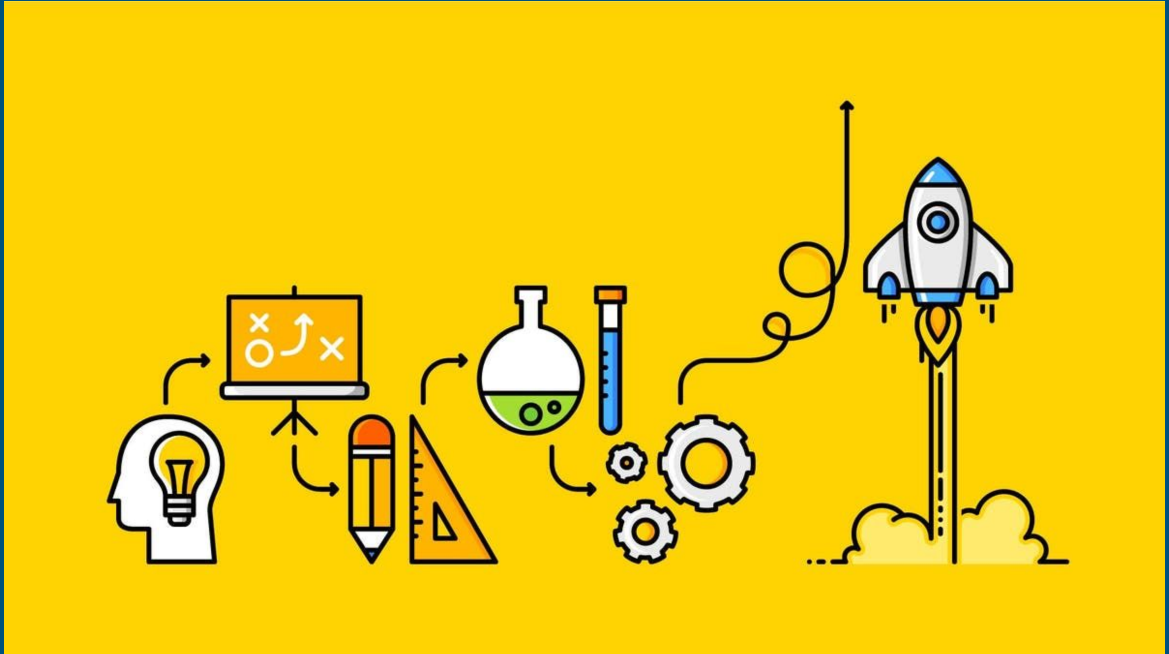
We ultimately selected the Nightingale Rose Chart to effectively represent the top and bottom airports.

How to Visualize? … Some ideas we had!

Design #3 focuses on uncovering trends and detecting patterns in airport delays over time.

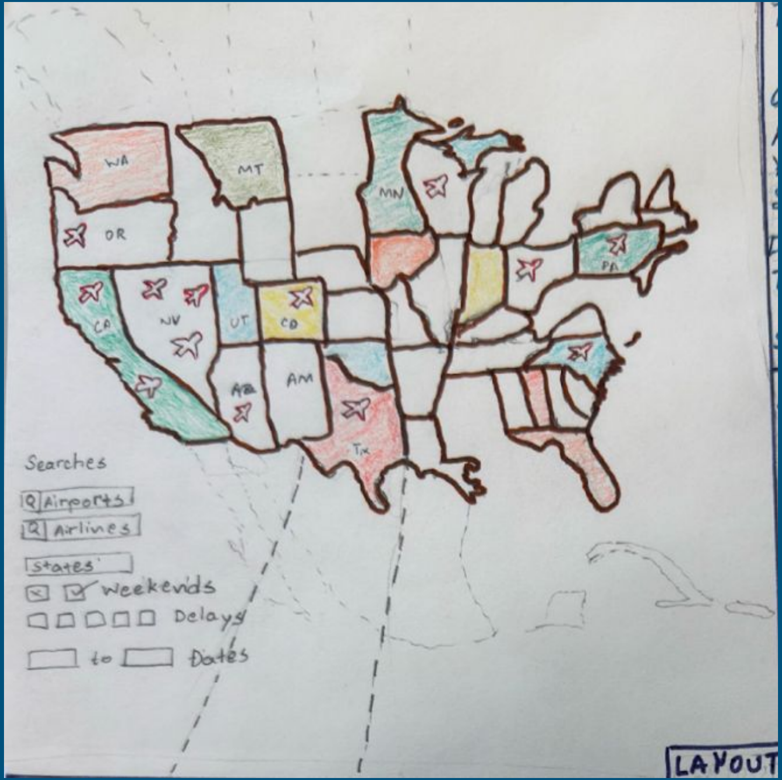
A key idea was to compare weekday and weekend delays to reveal insights into how flight activity and rush hours may impact delays.

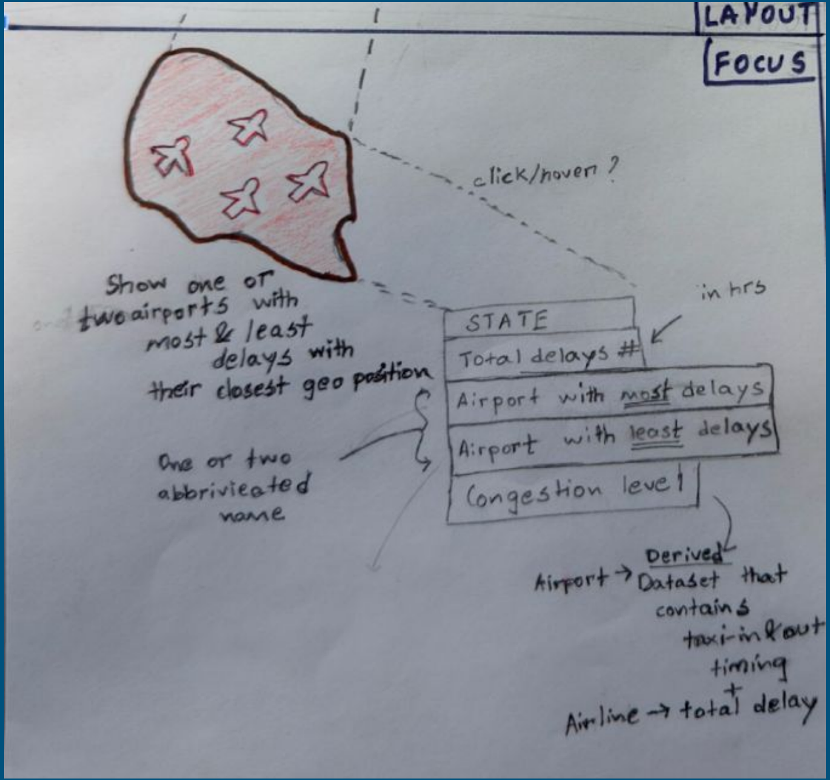




Design Evolution

Final Project Proposal - Design #1



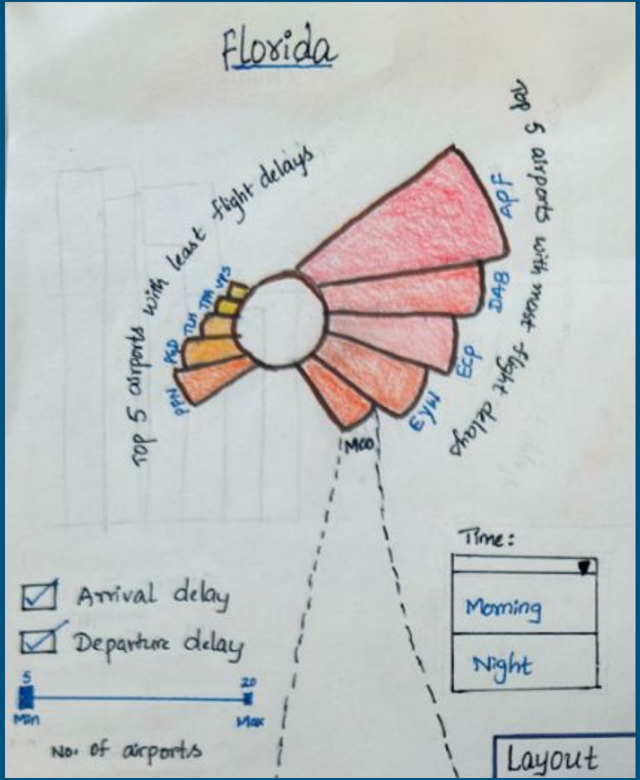


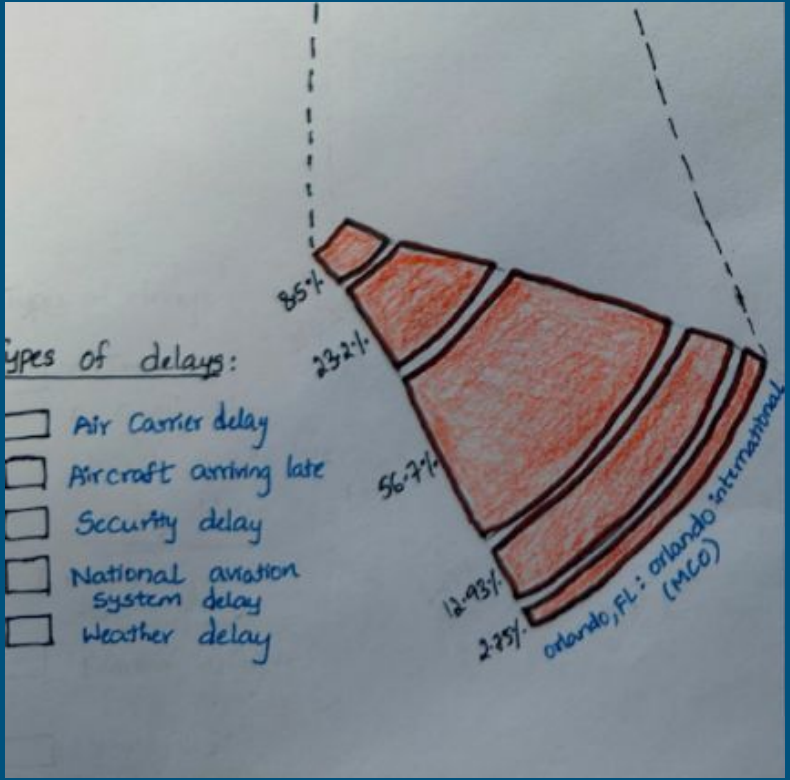
Final Project Proposal - Design # 1

Important things we are keeping in mind as we work on implementing this design:

- Popup placement (when you hover on a airport) should not collide with map elements
- The gradient key should be shown, so users understand what they’re seeing.
- Selection/filtering should grey out “irrelevant” regions (everything a user has not selected)
- Keep in mind considerations for visual impairments

Final Project Proposal - Design #2



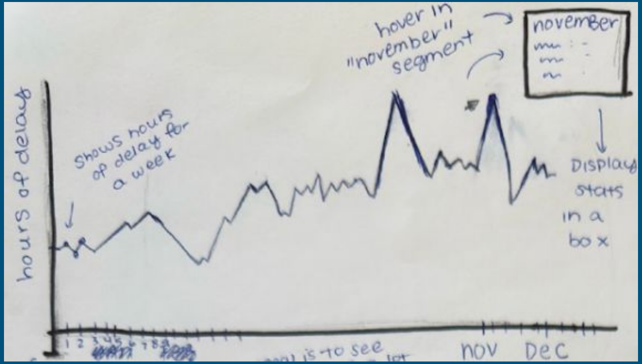


Final Project Proposal - Design #2

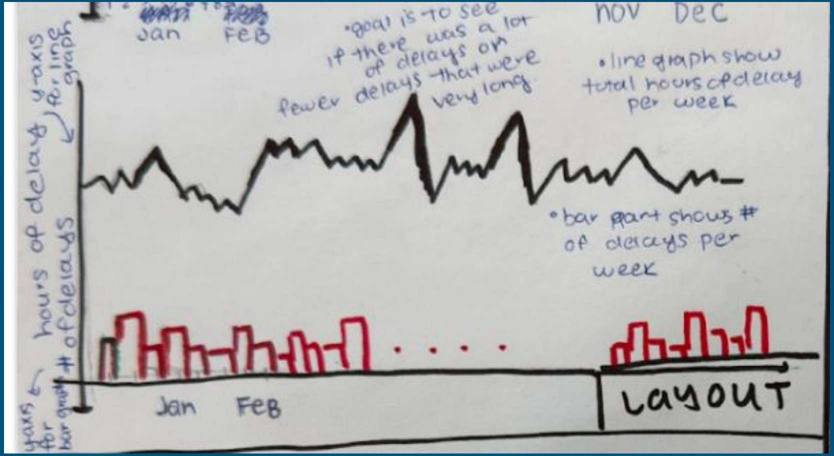
- Each airport, whether having the most flight delays or least flight delays, is represented as a segment around the circle.
- **NEW**: The Area of each segment corresponds to the total delays at that airport.
- Airports with highest/lowest number of flight delays are represented.
- When a user selects a segment (airport), a breakdown of different types of delays is displayed as a percentage.

Final Project Proposal - Design #3

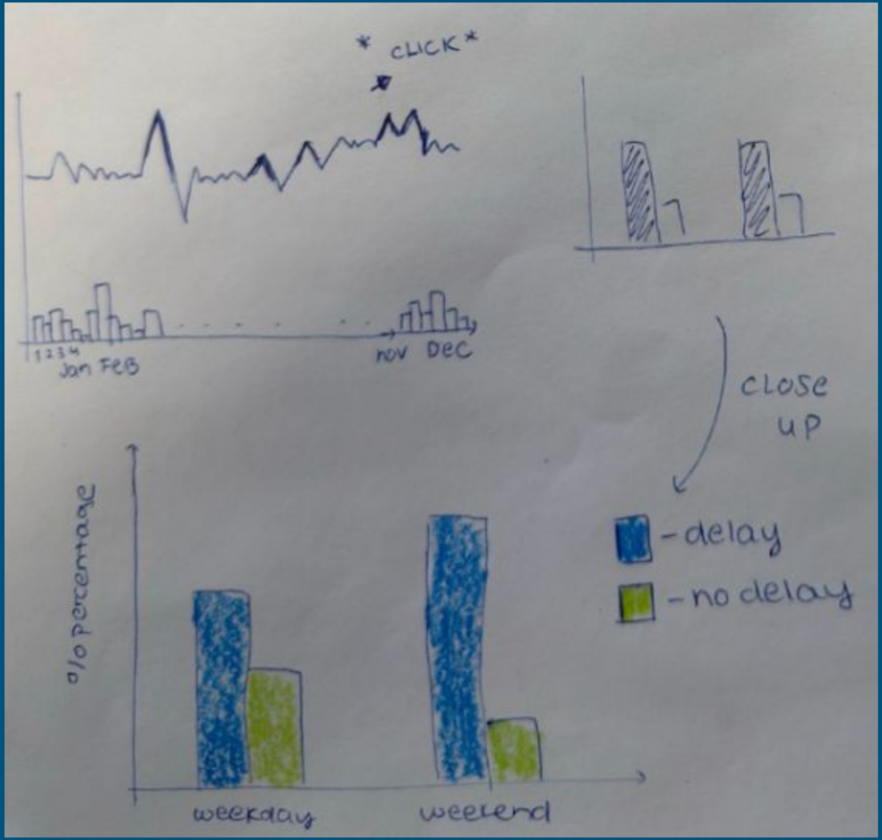
OLD



NEW



***CLICK**



Final Project Proposal - Design #3

NEW: 46 data points, 1 for each month between 2023 January - 2024 December.

The visualization graphs a subset of data points, segmented by month. The line gets darker/thicker as the # of delays (y-axis) gets higher.

When you click on a segment (**NEW**: segment “lines” - separated by month, but not visible/is relatively clear), shows a bar graph with weekend/weekday delay info (Goal: help user visualize).

Milestone 1: Data Pre-processing

Pre-Processing Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	YEAR	MONTH	DAY_OF_MO	OP_UNIQUE_	ORIGIN_AIRI	ORIGIN	ORIGIN_STA	ORIGIN_STA	DEST_AIRPO	DEST	DEST_CITY_N	DEST_STATE_	DEST_STATE_	CRS_DEP_TI	DEP_TIME	DEP_DELAY	DEP_DELAY_I	TAXI_OUT	WHEELS_OF	WHEELS_ON	TAXI_IN	CRS_ARR_TI	ARR_TIME	ARR_DELAY_I	CRS_ELAPSE	ACTUAL_ELA	AIR_TIME	DISTANCE	CARRIER_DE	WEATHER_D	NAS_DELAY	SECURITY_DE	LATE_AIRCRAFT_DELAY	
2	2023	1	1	9E	10135	ABE	PA	Pennsylvania	10397	ATL	Atlanta, GA	GA	Georgia	0600	0601	1	1	13	0614	0807	9	0825	0816	0	145	135	113	692						
3	2023	1	1	9E	10146	ABY	GA	Georgia	10397	ATL	Atlanta, GA	GA	Georgia	0736	0724	-12	0	24	0748	0822	22	0844	0844	0	68	80	34	145						
4	2023	1	1	9E	10146	ABY	GA	Georgia	10397	ATL	Atlanta, GA	GA	Georgia	1255	1245	-10	0	10	1255	1336	8	1355	1344	0	60	59	41	145						
5	2023	1	1	9E	10185	AEX	LA	Louisiana	10397	ATL	Atlanta, GA	GA	Georgia	1150	1855	425	425	24	1919	2132	8	1433	2140	427	103	105	73	500	0	83	2	0	342	
6	2023	1	1	9E	10185	AEX	LA	Louisiana	10397	ATL	Atlanta, GA	GA	Georgia	1748	1919	91	91	10	1929	2144	5	2037	2149	72	109	90	75	500	11	0	0	0	61	
7	2023	1	1	9E	10208	AGS	GA	Georgia	10397	ATL	Atlanta, GA	GA	Georgia	1745	1741	-4	0	9	1750	1819	14	1855	1833	0	70	52	29	143						
8	2023	1	1	9E	10257	ALB	NY	New York	11433	DTW	Detroit, MI	MI	Michigan	0545	1052	307	307	19	1111	1233	10	0755	1243	288	130	111	82	489	0	0	0	0	288	
9	2023	1	1	9E	10257	ALB	NY	New York	11433	DTW	Detroit, MI	MI	Michigan	1230	1301	31	31	19	1320	1442	8	1436	1450	14	126	109	82	489						
10	2023	1	1	9E	10257	ALB	NY	New York	12953	LGA	New York, NY	NY	New York	0825	0823	-2	0	15	0838	0904	7	0939	0911	0	74	48	26	136						
11	2023	1	1	9E	10397	ATL	GA	Georgia	10135	ABE	Allentown/B	PA	Pennsylvania	2156	2151	-5	0	14	2205	2344	4	2349	2348	0	113	117	99	692						
12	2023	1	1	9E	10397	ATL	GA	Georgia	10146	ABY	Albany, GA	GA	Georgia	1110	1105	-5	0	23	1128	1200	4	1210	1204	0	60	59	32	145						
13	2023	1	1	9E	10397	ATL	GA	Georgia	10146	ABY	Albany, GA	GA	Georgia	1920	1928	8	8	12	1940	2009	4	2016	2013	0	56	45	29	145						
14	2023	1	1	9E	10397	ATL	GA	Georgia	10185	AEX	Alexandria, L	LA	Louisiana	1010	1016	6	6	29	1045	1643	4	1105	1647		115			500						
15	2023	1	1	9E	10397	ATL	GA	Georgia	10185	AEX	Alexandria, L	LA	Louisiana	1620	1730	70	70	29	1759	1821	4	1708	1825	77	108	115	82	500	61	0	7	0	9	
16	2023	1	1	9E	10397	ATL	GA	Georgia	10208	AGS	Augusta, GA	GA	Georgia	1515	1511	-4	0	8	1519	1549	8	1613	1557	0	58	46	30	143						
17	2023	1	1	9E	10397	ATL	GA	Georgia	10208	AGS	Augusta, GA	GA	Georgia	2110	2106	-4	0	18	2124	2154	3	2210	2157	0	60	51	30	143						
18	2023	1	1	9E	10397	ATL	GA	Georgia	10731	BQK	Brunswick, G	GA	Georgia	0900	1436	336	336	9	1445	1526	3	1012	1529	317	72	53	41	238	12	0	0	0	305	
19	2023	1	1	9E	10397	ATL	GA	Georgia	10731	BQK	Brunswick, G	GA	Georgia	1710	1705	-5	0	11	1716	1757	5	1814	1802	0	64	57	41	238						
20	2023	1	1	9E	10397	ATL	GA	Georgia	10781	BTR	Baton Rouge,	LA	Louisiana	0855	0853	-2	0	15	0908	1201	3	0940	1204		105			448						
21	2023	1	1	9E	10397	ATL	GA	Georgia	10781	BTR	Baton Rouge,	LA	Louisiana	1430	1447	17	17	16	1503	1512	6	1510	1518	8	100	91	69	448						
22	2023	1	1	9E	10397	ATL	GA	Georgia	10980	CHA	Chattanooga,	TN	Tennessee	1340	1333	-7	0	13	1346	1412	4	1432	1416	0	52	43	26	106						
23	2023	1	1	9E	10397	ATL	GA	Georgia	10990	CHO	Charlottesvil	VA	Virginia	1300	1258	-2	0	12	1310	1419	2	1430	1421	0	90	83	69	457						
24	2023	1	1	9E	10397	ATL	GA	Georgia	10990	CHO	Charlottesvil	VA	Virginia	2055	2050	-5	0	15	2105	2209	4	2230	2213	0	95	83	64	457						
25	2023	1	1	9E	10397	ATL	GA	Georgia	11146	CRW	Charleston/D	WV	West Virginia	1100	1207	67	67	15	1222	1321	3	1225	1324	59	85	77	59	363	59	0	0	0	0	
26	2023	1	1	9E	10397	ATL	GA	Georgia	11146	CRW	Charleston/D	WV	West Virginia	2122	2124	2	2	12	2136	2225	7	2247	2232	0	85	68	49	363						
27	2023	1	1	9E	10397	ATL	GA	Georgia	11150	CSG	Columbus, G	GA	Georgia	1100	1119	19	19	20	1139	1204	4	1150	1208	18	50	49	25	83	0	18	0	0	0	
28	2023	1	1	9E	10397	ATL	GA	Georgia	11150	CSG	Columbus, G	GA	Georgia	2050	2040	-10	0	17	2057	2115	3	2139	2118	0	49	38	18	83						
29	2023	1	1	9E	10397	ATL	GA	Georgia	11193	CVG	Cincinnati, O	KY	Kentucky	1000	0958	-2	0	31	1029	1131	5	1122	1136	14	82	98	62	373						
30	2023	1	1	9E	10397	ATL	GA	Georgia	11308	DHN	Dothan, AL	AL	Alabama	0917	0947	30	30	14	1001	0936	3	0919	0939	20	62	52	35	170	20	0	0	0	0	
31	2023	1	1	9E	10397	ATL	GA	Georgia	11308	DHN	Dothan, AL	AL	Alabama	1657	1652	-5	0	13	1705	1639	6	1658	1645	0	61	53	34	170						
32	2023	1	1	9E	10397	ATL	GA	Georgia	11423	DSM	Des Moines, I	IA	Iowa	1105	1103	-2	0	24	1127	1211	5	1228	1216	0	143	133	104	743						
33	2023	1	1	9E	10397	ATL	GA	Georgia	11612	EVV	Evansville, IN	IN	Indiana	1350	1344	-6	0	18	1402	1403	5	1411	1408	0	81	84	61	350						
34	2023	1	1	9E	10397	ATL	GA	Georgia	11612	EVV	Evansville, IN	IN	Indiana	2135	0810	635	635	9	0819	0819	3	2158	0822	624	83	72	60	350	0	624	0	0	0	
35	2023	1	1	9E	10397	ATL	GA	Georgia	11624	EYW	Key West, FL	FL	Florida	1740	1812	32	32	16	1828	1954	3	1933	1957	24	113	105	86	646	0	0	0	0	24	
36	2023	1	1	9E	10397	ATL	GA	Georgia	11641	FAY	Fayetteville, N	NC	North Carolir	1354	1410	16	16	12	1422	1510	4	1512	1514	2	78	64	48	331						
37	2023	1	1	9E	10397	ATL	GA	Georgia	11953	GNV	Gainesville, F	FL	Florida	0955	0953	-2	0	24	1017	1110	4	1120	1114	0	85	81	53	300						
38	2023	1	1	9E	10397	ATL	GA	Georgia	11953	GNV	Gainesville, F	FL	Florida	1510	1508	-2	0	12	1520	1606	4	1624	1610	0	74	62	46	300						
39	2023	1	1	9E	10397	ATL	GA	Georgia	11973	GPT	Gulfport/Bilc	MS	Mississippi	1005	1005	0	0	30	1035	1031	8	1035	1039	4	90	94	56	352						
40	2023	1	1	9E	10397	ATL	GA	Georgia	12007	GTR	Columbus, M	MS	Mississippi	1157	1250	53	53	17	1307	1252	3	1206	1255	49	69	65	45	241	49	0	0	0	0	
41	2023	1	1	9E	10397	ATL	GA	Georgia	12007	GTR	Columbus, M	MS	Mississippi	2100	2055	-5	0	14	2109	2053	4	2118	2057	0	78	62	44	241						
42	2023	1	1	9E	10397	ATL	GA	Georgia	12197	HPN	White Plains,	NY	New York	0950	0947	-3	0	29	1016	1209	4	1214	1213	0	144	146	113	780						
43	2023	1	1	9E	10397	ATL	GA	Georgia	12323	ILM	Wilmington,	NC	North Carolir	0900	0856	-4	0	26	0922	1019	4	1023	1023	0	83	87	57	377						
44	2023	1	1	9E	10397	ATL	GA	Georgia	12323	ILM	Wilmington,	NC	North Carolir	1315	1405	50	50	47	1452	1546	3	1439	1549	70	84	104	54	377	50	0	20	0	0	
45	2023	1	1	9E	10397	ATL	GA	Georgia	12448	JAN	Jackson/Vick	MS	Mississippi	2105	2103	-2	0	13	2116	2111	6	2136	2117	0	91	74	55	341						
46	2023	1	1	9E	10397	ATL	GA	Georgia	12945	LEX	Lexington, KY	KY	Kentucky	1000	1016	16	16	109	1205	1258	7	1120	1305	105	80	169	53	304	16	0	89	0	0	
47	2023	1	1	9E	10397	ATL	GA	Georgia	12945	LEX	Lexington, KY	KY	Kentucky	1559	1554	-5	0	16	1610	1701	5	1715	1706	0	76	72	51	304						
48	2023	1	1	9E	10397	ATL	GA	Georgia	13277	MGM	Montgomery,	AL	Alabama	0825	0841	16	16	12	0853	0827	4	0829	0831	2	64	50	34	147						
49	2023	1	1	9E	10397	ATL	GA	Georgia	13277	MGM	Montgomery,	AL	Alabama	1200	1158	-2	0	17	1215	1146	3	1159	1149	0	59	51	31	147						
50	2023	1	1	9E	10397	ATL	GA	Georgia	13277	MGM	Montgomery,	AL	Alabama	1721	1712	-9	0	13	1725	1654	4	1717	1658	0	56	46	29	147						
51	2023	1	1	9E	10397	ATL	GA	Georgia	13377	MLU	Monroe, LA	LA	Louisiana	1410	1412	2	2	28	1440	1455	4	1455	1459	4	105	107	75	448						
52	2023	1	1	9E	10397	ATL	GA	Georgia	13377	MLU	Monroe, LA	LA	Louisiana	1930	1926	-4	0	12	1938	1950	5	2015	1955	0	105	89	72	448						
53	2023	1	1	9E	10397	ATL	GA	Georgia	13422	MOB	Mobile, AL	AL	Alabama	0839	0829	-10	0	19	0848	0846	4	0906	0850	0	87	81	58	302						
54	2023	1	1	9E	10397	ATL	GA	Georgia	13422	MOB	Mobile, AL	AL	Alabama	1340	1349	9	9	17	1406	1356	3	1400	1359	0	80	70	50	302						
55	2023	1	1	9E	10397	ATL	GA	Georgia	13422	MOB	Mobile, AL	AL	Alabama	2141	2143	2	2	16	2159	2151	5	2205	2156	0	84	73	52	302						
56	2023	1	1	9E	10397	ATL	GA	Georgia	13495	MSY	New Orleans,	LA	Louisiana	1045	1040	-5	0	24	1104	1114	3	1125	1117	0	100	97	70	425						
57	2023	1	1	9E	10397	ATL	GA	Georgia	14574	ROA	Roanoke, VA	VA	Virginia	1040	1036	-4	0	19	1055	1157	6	1211	1203	0	91	87	62	357						
58	2023	1	1	9E	10397	ATL	GA	Georgia	14783	SGF	Springfield, I	MO	Missouri	1130	1125	-5	0	17	1142	1210	4	1225	1214	0	115	109	88	563						

The data was obtained from a public source in monthly CSV chunks.

Our first step was to combine all the individual files into a single consolidated dataset.

This was accomplished using a Python preprocessing script that streamlined the next analysis and calculations for our visual designs.

03/05 (Yash)

Processing Data for Geo Map

Structured the data by grouping on following controls,

- Year-Month
- State
- Airport
- Airline
- IsWeekday

For each group, taking average of,

- 5 delay types values
- Congestion (derived)
- Time deviation
- Speed deviation
- Recovery efficiency

We do this for both arrival and departure delays

2023-1	AK	Adak Island, AK: Adak	Alaska Airlines Inc.	False	0	0	25	39	0	0.115584	0.043436	-0.036366	2.548913
		Anchorage, AK: Ted Stevens Anchorage International	Alaska Airlines Inc.	False	415	0	781	0	346	0.197400	0.087104	-0.066561	1.186857
				True	1883	98	2131	0	1250	0.183852	0.080514	-0.064808	1.669687
			American Airlines Inc.	True	494	0	25	0	130	0.111392	0.066489	-0.059524	0.040064
			Delta Air Lines Inc.	False	367	71	133	0	154	0.152717	0.041582	-0.033095	1.055420
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2024-9	WY	Riverton/Lander, WY: Central Wyoming Regional	SkyWest Airlines Inc.	True	376	5	59	0	321	0.442222	0.182804	-0.103294	2.891755
		Rock Springs, WY: Southwest Wyoming Regional	SkyWest Airlines Inc.	False	77	0	3	0	0	0.444444	0.069767	-0.040000	0.038961
				True	238	0	26	0	125	0.430556	0.079018	-0.035855	0.003402
		Sheridan, WY: Sheridan County	SkyWest Airlines Inc.	False	426	0	0	0	0	0.375000	0.098631	-0.027512	0.482100
				True	403	0	17	0	0	0.329545	0.025039	-0.008123	-0.024050

82174 rows × 9 columns

03/10 (Yash)

Processing Data for Airports

	Adak Island, AK: Adak	2	62	48	153	417	1255
AK	Anchorage, AK: Ted Stevens Anchorage International	0	20066	2194	8551	112	13182
		1	46023	17089	18917	2	21953
		2	51432	9593	23404	126	60688
		3	67601	1425	21930	117	54718
...	...	...	...	...	...	...	...
WY	Riverton/Lander, WY: Central Wyoming Regional	2	3334	497	1219	0	2548
		3	0	0	0	0	46
	Rock Springs, WY: Southwest Wyoming Regional	1	7894	602	767	0	2127
		2	2751	490	765	0	1588
	Sheridan, WY: Sheridan County	1	8349	1704	963	4	4887

1054 rows × 5 columns

Structured the data by grouping on following controls,

- State
- Airport
- Departure time slots (derived)

For each group, taking average of,

- 5 delay type values

We do this for both arrival and departure delays

03/14 (Sreevidya)

Processing Data for Trends

2023-1	**False**	17926	1600	15023	212	13173	816998	110635	327826	4851	677415
	True	48996	5449	48997	424	44482	2239344	455288	1496820	12434	2599003
2023-10	**False**	17759	834	12815	179	15101	800227	54105	273550	4576	786557
	True	39881	2196	29955	377	34632	1688595	169396	592063	9174	1722730
2023-11	**False**	14088	773	11899	182	11874	611963	57185	251043	4292	614201
	True	32746	1821	27945	452	26270	1424486	129232	520815	9538	1252892
2023-12	**False**	17710	1677	16383	266	15664	777284	161787	370687	7367	880239
	True	34894	2020	30068	497	29960	1434672	232889	665890	11622	1540525
2023-2	**False**	15798	793	10839	167	11506	740413	60494	229056	5609	646875
	True	40102	4357	37566	282	34836	1820531	331944	930465	8833	1880013
2023-3	**False**	23674	1537	21156	303	21779	973430	106204	569799	8165	1159306
	True	54621	4867	52445	483	49774	2226673	339888	1337707	10580	2503455
2023-4	**False**	26708	2468	22501	310	24047	1182952	183412	703827	9641	1380688
	True	48994	4378	43130	435	45637	2025004	306990	1122716	9710	2386687
2023-5	**False**	16213	946	11218	177	12782	721679	60965	260901	5530	702018
	True	49740	3975	38527	434	43866	2076438	262934	903497	11688	2304147

Structured the data by grouping on following controls,

- Year-Month
- IsWeekday

For each group taking,

- average of 5 delay type values
- count of 5 delay types

We will use the average value to draw a trend line while counts for the volume (bars) on the same graph

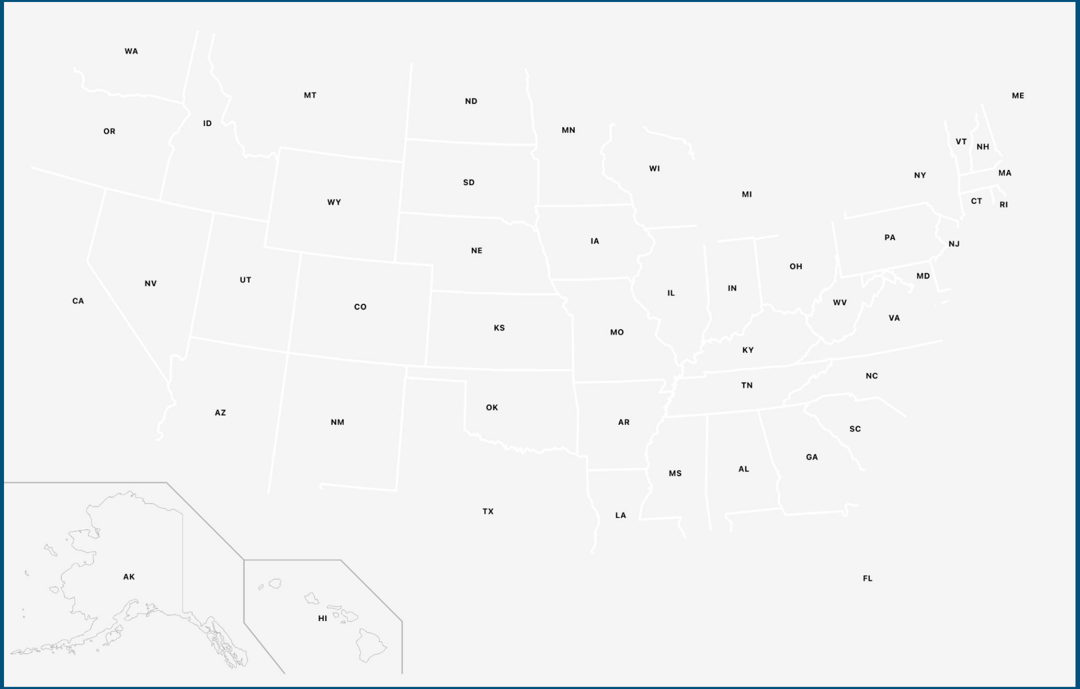
03/19 (Aditi)

Milestone 2: Visualization
GEOMAP

Designing Geo Map

As a starting point for visualizing airport delays across the U.S., I created a map of the United States with svg labeled with state names.

At this point, the focus is on scaling, sizing and positioning of the texts and Map on the window for the next steps.



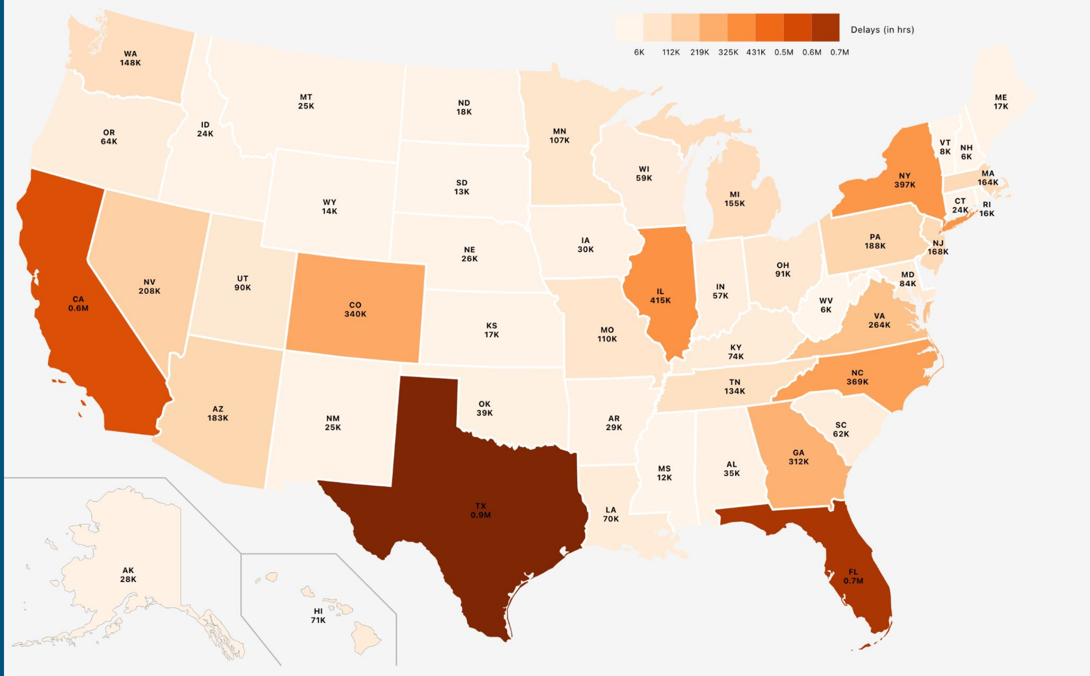
04/01 (Yash)

Designing Geo Map

Building on the labeled U.S. map, I completed the shading of each state based on the average airport delays.

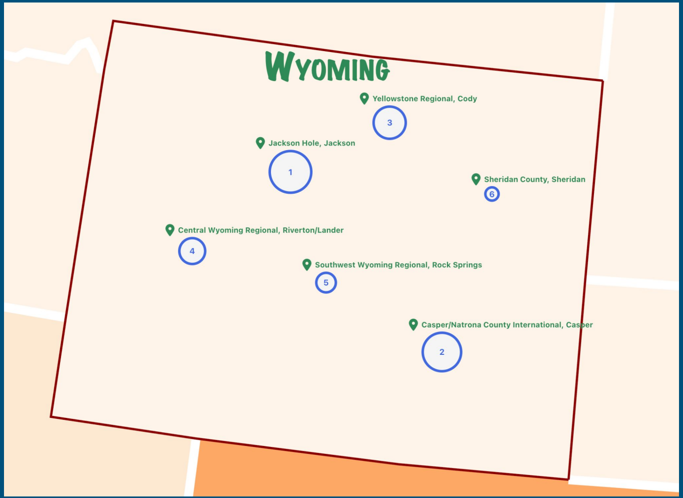
I also added the actual delay values directly onto the states to provide precise, at-a-glance information.

To support interpretation, I included a sequential legend that maps the color scale to delay durations, making the visualization both intuitive and informative.



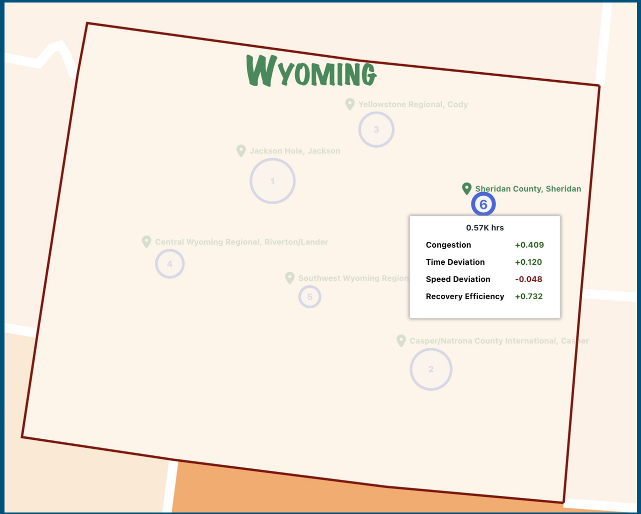
04/03 (Yash)

Geo Map - Zoomed



Adding the hover feature for each airport marker. When a user hovers over an airport, it gets visually highlighted and a popup appears showing detailed information, including the total delay and key derived metrics.

To help users quickly interpret performance, I incorporated visual cues—using green to indicate improvements or lower delays, and red for increases or higher delays. This immediate color feedback, combined with contextual data, allows users to quickly assess the situation at each airport.

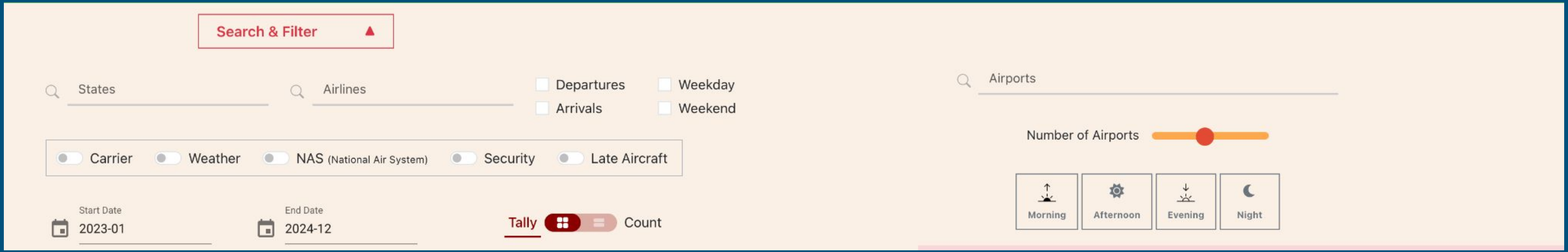


To enhance interactivity, implemented D3 zooming functionality, allowing users to click on a state and zoom in for a closer view. As the map zooms, dynamically adjust the text sizes to maintain readability at different scales.

I added airport locations within each state using accurate (x, y) coordinates sourced online. For each airport, I placed a location marker and overlaid a circle whose size represents the corresponding delay value. This layered approach provides a more detailed and interactive way to explore delay patterns at both the country and state levels.

04/03 (Yash)

Geo Map Controls



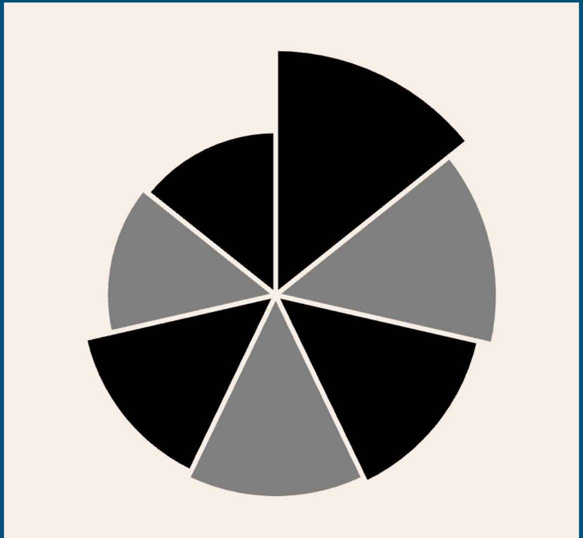
Designed and integrated interactive controls for both the geo map and airport-level data. Based on the designs we outlined during the planning stage, I added input elements such as dropdowns, switches, dates etc. that allow users to filter and adjust the visualization dynamically.

When these inputs are changed, the map updates in real time to reflect the new delay distributions, enabling a more tailored and insightful exploration of the data. The goal was to keep the interaction intuitive while still offering powerful filtering options.

04/07 (Yash)

Milestone 3: Visualization Airports

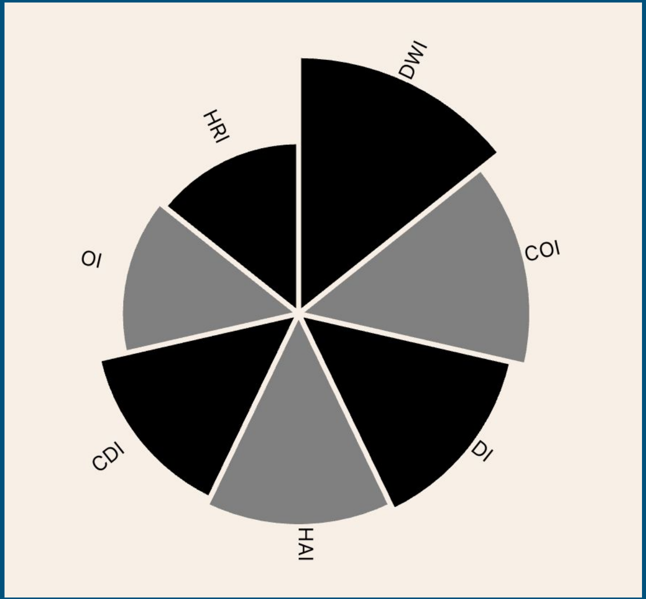
Designing Airport delays



I integrated a Nightingale Rose Chart to represent the proportion of total delays attributed to each airport. Each segment of the chart corresponds to a specific airport, with the size of the segment directly proportional to its cumulative delay time.

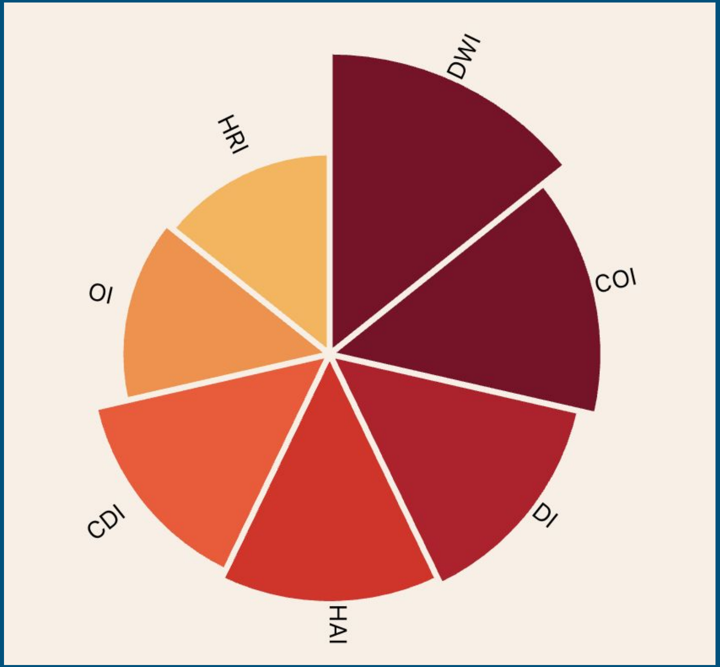
When a user hovers over any segment, the corresponding slice slightly expands to emphasize the selection and provide immediate visual feedback.

Adding the abbreviated airport names along the arc. When hovered, they enlarge.



03/25 (Sreevidya)

Designing Airport Delays



Refined the Nightingale Rose Chart by introducing color hue as an additional encoding channel alongside segment length. This dual encoding allows users to distinguish airports not just by the magnitude of their delays (indicated by arc length), but also by the intensity of the color applied to each segment.

The arcs are designed with equal angular widths, creating a radial bar chart where only color and radial length vary—ensuring consistency while emphasizing delay differences more effectively.

On initial loading, top seven airports are shown from all the states. This provides immediate, high-impact insight into the most problematic hubs.

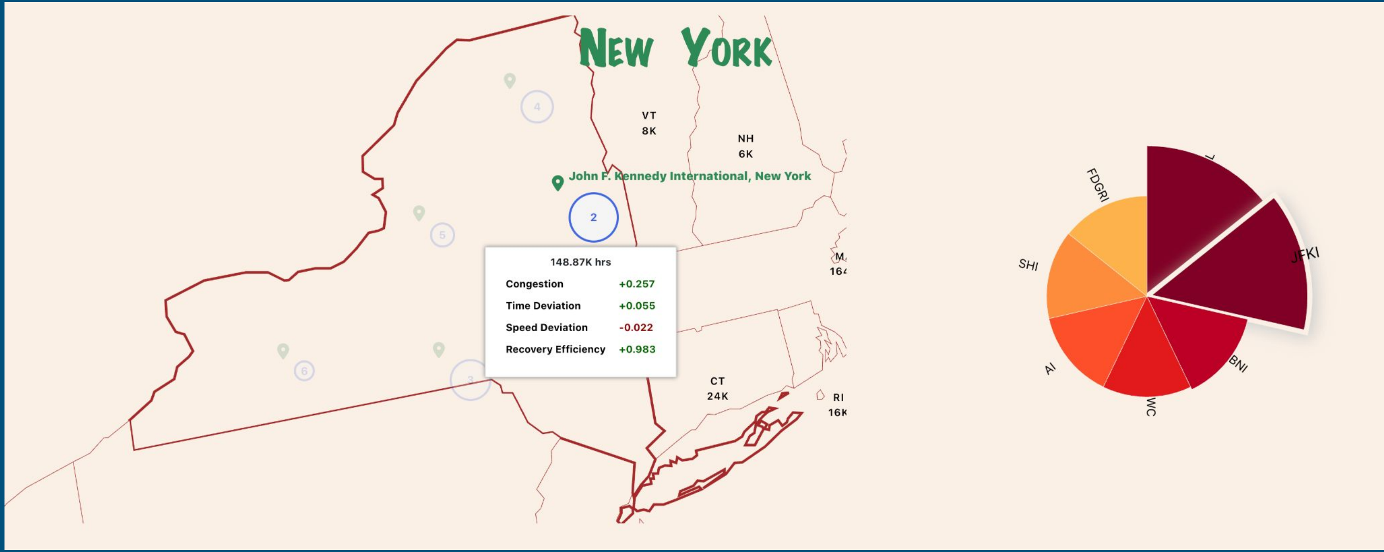
On hover, the segmented pie will enlarge to give a visual feedback.

04/03 (Sreevidya)

Designing Airport Delays

Integrated a linked view with the Geo Map: when a user clicks on a specific state, the Nightingale Rose Chart dynamically updates to display only the airports located within that selected state. This interaction allows for targeted analysis of regional delay patterns, supporting both national overviews and localized deep dives.

Also, when a user hovers over a specific airport segment, that segment expands slightly for emphasis and dynamically displays detailed statistics about the selected airport in the map view.

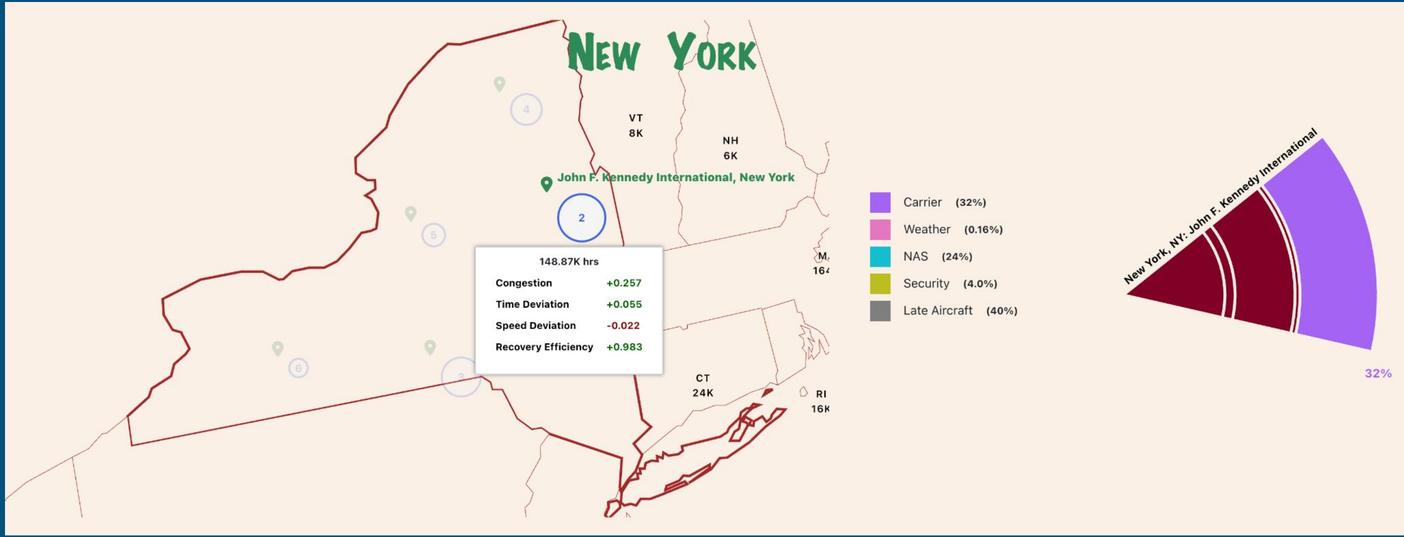


04/10 (Sreevidya)

Designing Airport Delays

Developed a drill-down mechanism within the Nightingale Rose Chart interface. This enhancement enables users to explore delay types at an individual airport level, offering a segmented and intuitive breakdown of contributing factors. Upon clicking a specific airport segment in the main Nightingale Rose Chart, the view transitions into a segmented radial chart dedicated to that airport. This secondary chart decomposes the total delay into its five standard FAA delay categories.

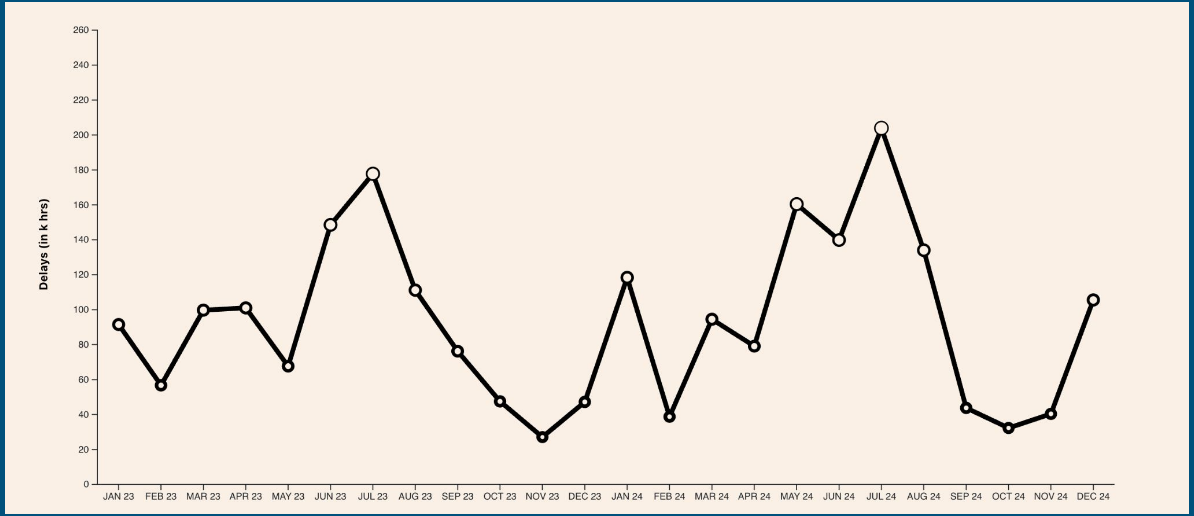
Hovering over any delay type segment changes its color dynamically and reveals a tooltip displaying the exact percentage that delay type contributes to the total delays at the selected airport. A color-coded legend is persistently displayed on the side, showing all five delay categories and their corresponding percentages, reinforcing the link between color and meaning while supporting users with accessibility needs.



04/17 (Sreevidya)

Milestone 4: Visualization Trends

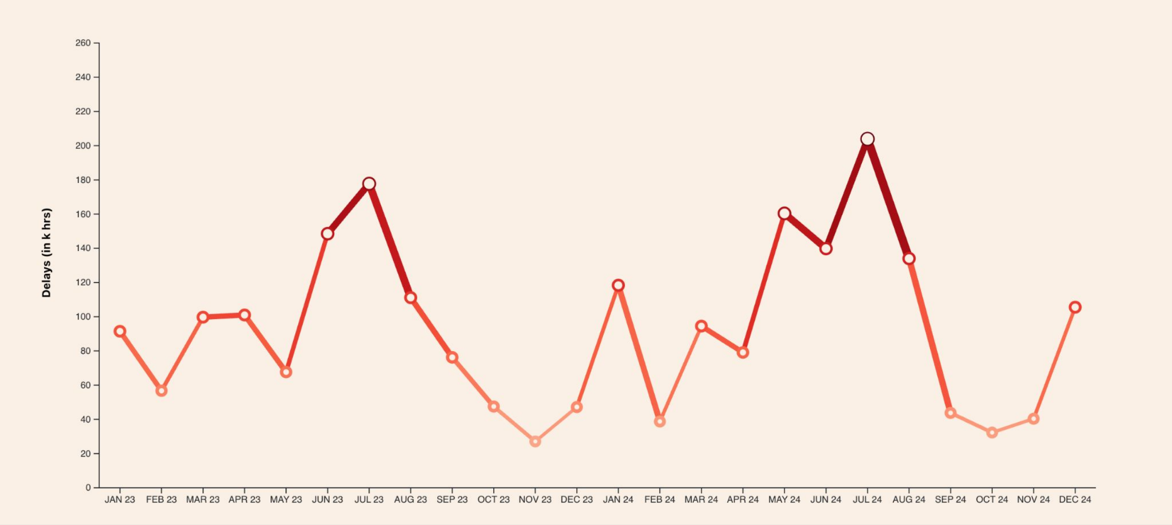
Designing Delay Trends



Developed a basic line chart that visualizes total delays (aggregated in thousands of hours) across a rolling 24-month period. Each data point on the line represents the total delay for a single month, offering a clear and continuous view of how delay volumes evolve over time. Each month has a point (circle) that will be used for further interactions.

03/31 (Aditi)

Designing Delay Trends

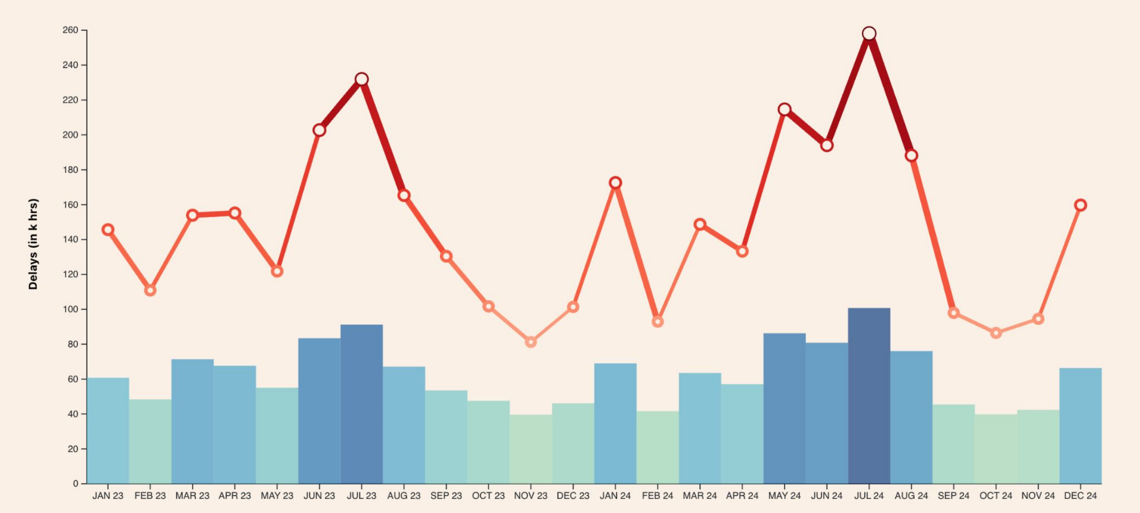


To improve the expressiveness and visual impact of the monthly delay trend line chart, I introduced multi-channel visual encoding using both line thickness and color intensity to represent the magnitude of delays for each month.

The width of the line segment between two months now varies proportionally with the delay magnitude. The color intensity of the line changes along the time axis. Darker or more saturated hues indicate months with higher delay volumes

04/04 (Aditi)

Designing Delay Trends

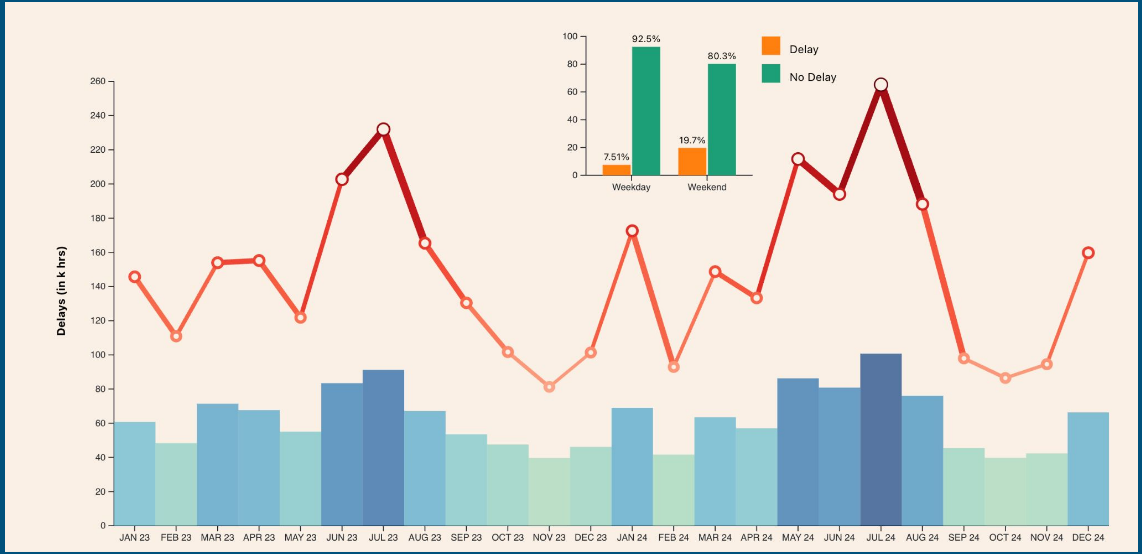


Adding a dual-axis bar chart that visualizes the number of individual delays occurring each month.

To add depth to the visual encoding, the bars are color-coded using intensity gradients: darker or more saturated bars indicate months with a higher volume of delay incidents.

04/11 (Aditi)

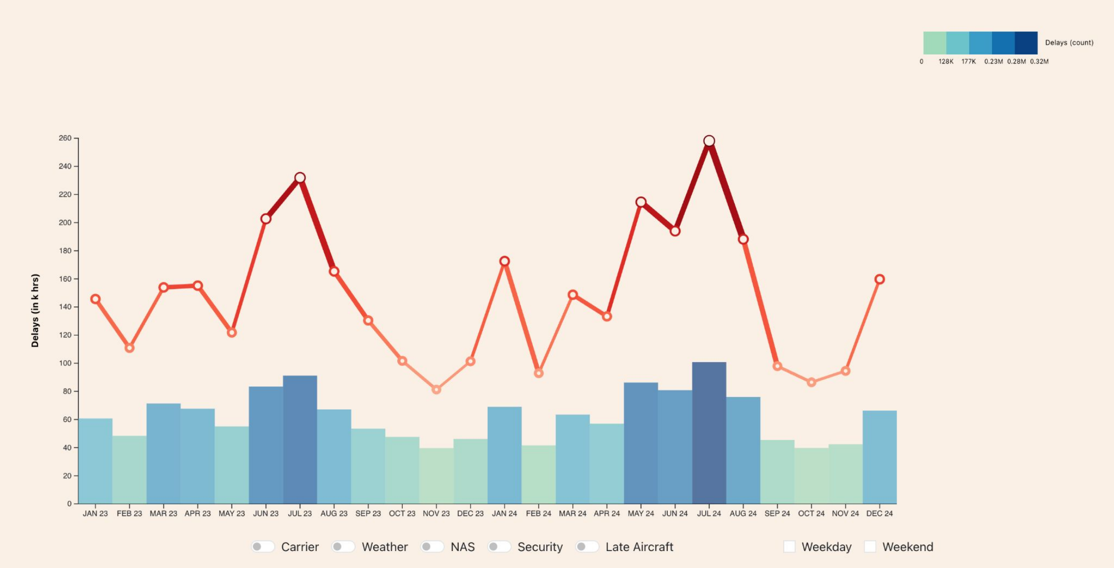
Designing Delay Trends



To provide actionable insights into the behavioral patterns behind monthly delay trends, implemented an interactive drill-down feature tied to the primary time-series line and bar charts. When a user clicks on any month's data point, a dedicated grouped bar chart is rendered to display a comparative breakdown of delayed vs. non-delayed flights, segmented by weekday vs. weekend for the selected month.

04/16 (Aditi)

Designing Delay Trends



Adding a persistent legend adjacent to the chart to map color hues to their corresponding to number of delays.

A control panel allows users to customize the data view in real-time, providing filtering capabilities triggered by the control panel initiates a smooth animated transition.

04/18 (Aditi)

Evaluation

Key Insights from Visualizations

Our visualizations uncovered several important patterns and insights about U.S. airport delays:

★ **Unexpected Cause Distribution**

- ❖ Carrier issues and late aircraft were primary delay sources nationwide
- ❖ Security delays consistently ranked as least common across all airports
- ❖ Weather delays were not as common as we originally anticipated - though there are seasonal patterns.

★ **Temporal Patterns**

- ❖ Morning flights showed highest delay frequency, contrary to evening congestion hypothesis
- ❖ Weekdays experienced nearly double the total hours of delay & # of delays as compared to weekends

★ **Geographic Discrepancies**

- ❖ Texas: higher total delay durations
- ❖ California: higher frequency of delay events
- ❖ Suggests different underlying causes (operational issues vs. congestion)

Key Insights from Visualizations (Contd.)

★ **<u>Correlation Patterns</u>**

❖ Delay frequency and severity rise and fall together without significant outliers

★ **<u>Performance Variations</u>**

❖ Clear differences between airports handling similar volumes
❖ Some facilities showed better recovery efficiency despite similar congestion

★ **<u>System Interdependencies</u>**

❖ Hub airport delays created cascading effects throughout connected destinations
❖ Network-level perspective revealed through multi-level exploration

How We Answered Our Research Questions

Building on these discoveries, our visualizations provided targeted answers to each of our initial research questions through specific interactive features and design elements:

★ **Question: How do flight delays vary geographically across the U.S.?**

- ❖ **Interactive map visualization** with color-encoded severity showed uneven distribution
- ❖ Texas showed higher total delay durations while California had higher frequency
- ❖ **State and airport drill-down** capabilities identified regional patterns
- ❖ **Filtering features** allowed isolation of patterns by carrier and region

★ **Question: What factors contribute most significantly to flight delays?**

- ❖ **Nightingale Rose charts** revealed carrier issues and late aircraft as primary causes
- ❖ **Weather delays** showed seasonal patterns but were less significant than expected
- ❖ **Security delays** consistently ranked lowest across all airports
- ❖ **Proportional visualization** made previously hidden patterns immediately apparent

★ **Question: When do delays most frequently occur?**

- ❖ **Time-based filtering** revealed morning flights experience highest delay frequency
- ❖ **Weekday/weekend comparison** showed weekdays have nearly double the amount & # of delays
- ❖ **Monthly trends** captured seasonal variations, especially in weather-related delays
- ❖ **Time-of-day filters** enabled pattern isolation across different airports

How We Answered Our Research Questions (Contd.)

- ★ **Question: How do airports compare in handling similar challenges?**
  - ❖ **Airport cards with multiple metrics** (congestion, recovery efficiency, time and speed deviations) enabled benchmarking
  - ❖ **Side-by-side visualization** of multiple airports facilitated direct comparison
  - ❖ **Several facilities** demonstrated better recovery efficiency despite similar congestion
  - ❖ **Performance indicators** transformed raw delay data into actionable insights
- ★ **Question: Is there a correlation between delay frequency and severity?**
  - ❖ **Dual-axis trend visualization** confirmed strong correlation between hours and counts
  - ❖ **No significant outliers** appeared between delay volume and duration
  - ❖ **Delay frequency and severity** consistently rise and fall together
  - ❖ **Findings validated** our approach to measuring delay impact

During implementation, we encountered an unexpected challenge when trying to balance information density with clarity. The rich airport performance metrics (congestion, recovery efficiency, speed/time deviation) added analytical depth but risked overwhelming users. This led us to redesign the airport cards with progressive disclosure, showing summary metrics initially and revealing detailed breakdowns only on demand.

How Well Our Visualizations Work

Strengths:

★ **Multi-level geographic exploration** allows users to seamlessly navigate from country-wide patterns to individual airport metrics.

★ **Rich filtering system** enables precise analysis through state/airline selection, weekday/weekend toggles and delay type filters.

★ **Interactive elements** including hover states, tooltips and dynamic updates provide immediate feedback during exploration.

★ **Comparative metrics** for airports (congestion, recovery efficiency, time/speed deviation) provide actionable insights about operational performance.

★ **Coordinated views** between the map and detailed charts create connections between geographic patterns and causal factors.

How Well Our Visualizations Work

Strengths:

★ **Clear visual encoding** through color intensity and segment size helps users quickly identify patterns and hotspots.

★ **User-driven exploration** supports both guided analysis and open-ended discovery of delay patterns.

★ **Actionable insights** reveal not just where and when delays occur, but also underlying causes.

Informal user testing with peers revealed that while the visualizations analytical capabilities were appreciated, the state-level drill-down functionality wasn't immediately obvious to first-time users. This feedback influenced our decision to add more explicit visual cues for clickable elements and helped us identify areas where additional guidance would be beneficial.

How Well Our Visualizations Work (Contd.)

Limitations:

★ **Data Limitations:**

- ❖ **Temporal constraints** - Limited to 2023-2024 data, preventing multi-year historical analysis
- ❖ **Monthly aggregation** - May obscure important daily or hourly delay patterns; gave for a more broad scope for analysis
- ❖ **Domestic focus** - Analysis excludes international flights and carriers
- ❖ **External factors** - Lacks integration with weather data, passenger volume or other contextual datasets

★ **Technical Limitations:**

- ❖ **Limited mobile responsiveness** - Fixed SVG dimensions optimize for desktop viewing only
- ❖ **Performance challenges** - Complex visualizations cause slowdowns with large data selections
- ❖ **Browser compatibility** - Modern JavaScript features may create inconsistent experiences
- ❖ **Accessibility gaps** - Insufficient support for screen readers and keyboard navigation

How Well Our Visualizations Work (Contd.)

Limitations:

★ **User Experience Limitations**:

- ❖ **Complex interface** - Multiple controls create steep learning curve for new users
- ❖ **Limited guidance** - No built-in tutorial or onboarding for first-time visitors
- ❖ **Information density** - Rich data display may overwhelm users and obscure key insights
- ❖ **Search constraints** - Functionality limited to predefined categories

★ **Analytical Limitations**:

- ❖ **Descriptive focus** - Shows historical patterns without predictive capabilities
- ❖ **Limited comparison tools** - Insufficient mechanisms for direct airport/time period comparisons
- ❖ **Causal analysis gaps** - Visualizes delay types without explaining underlying causes
- ❖ **Contextual understanding** - Limited connection to broader transportation system issues

A particularly challenging implementation obstacle was managing the rendering performance of the Nightingale Rose Chart when displaying multiple airports simultaneously. Our initial design attempted to show too many airports at once, causing significant lag. This technical constraint shaped our final design decision to limit the default view to the top seven airports, with user controls to adjust this number based on their hardware capabilities.

Potential Improvements

- ★ <u>Data Enhancements</u>:
  - ❖ **Expand temporal range** beyond 2023-2024 for multi-year trend analysis
  - ❖ **Increase granularity** from monthly to hourly/daily delay patterns
  - ❖ **Include international flights** to provide global context
  - ❖ **Integrate external datasets** (weather, passenger volume, construction events)

- ★ <u>Technical Improvements</u>:
  - ❖ **Implement responsive design** for mobile/tablet compatibility
  - ❖ **Optimize performance** for large data selections and complex filtering
  - ❖ **Enhance accessibility** with ARIA attributes and screen reader support
  - ❖ **Improve cross-browser compatibility** across platforms

Potential Improvements (Contd.)

★ **User Experience:**

- ❖ **Create guided onboarding** for first-time users
- ❖ **Add contextual explanations** for technical metrics and visualizations
- ❖ **Simplify interface** with logical control grouping and progressive disclosure
- ❖ **Enable personalization** with saved views and preferences

★ **Advanced Capabilities:**

- ❖ **Develop predictive modeling** to forecast likely delay patterns
- ❖ **Enhance comparative analysis** with side-by-side views of airports/time periods
- ❖ **Create cause-effect tools** to better understand delay relationships
- ❖ **Add data export options** and shareable insights for collaboration

User feedback suggested that the technical terminology used for airport metrics (like "recovery efficiency") was difficult to understand without context. In future iterations, we would implement contextual help systems that explain these metrics directly within the interface, making the visualization more accessible to non-specialists while maintaining its analytical depth.

AIRPORT DELAY ANALYSIS: THE END