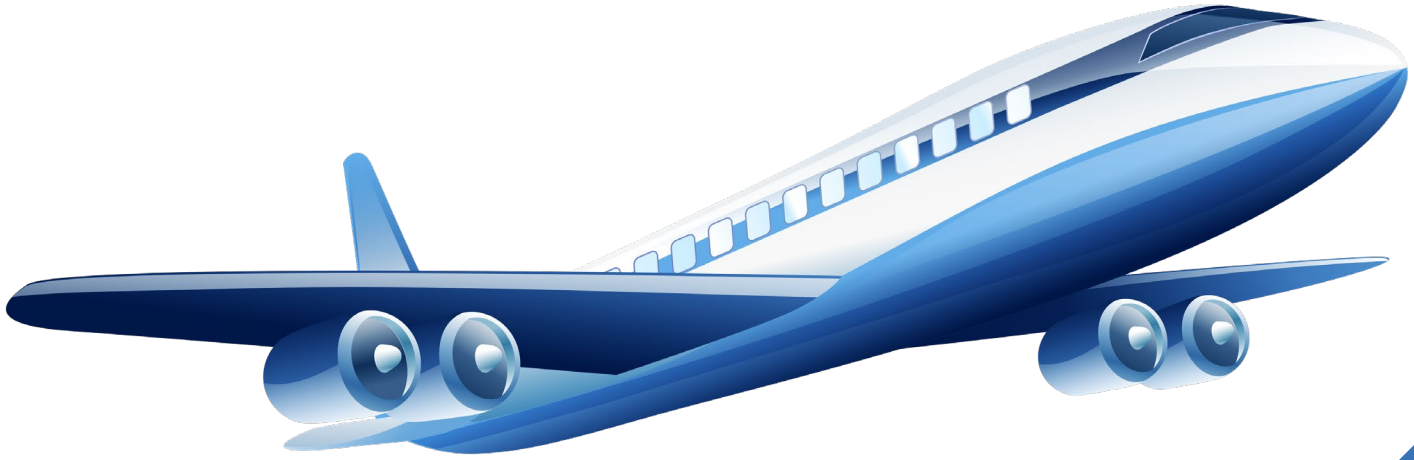
Taking Off:
Investigating Flight Delays



Blue Group 4:

Sarah Arnold, Sterling Hayden,

Marie Bennett, Rohan Venkatraman,

Fred Lindsey

Insights

46% of flights were delayed

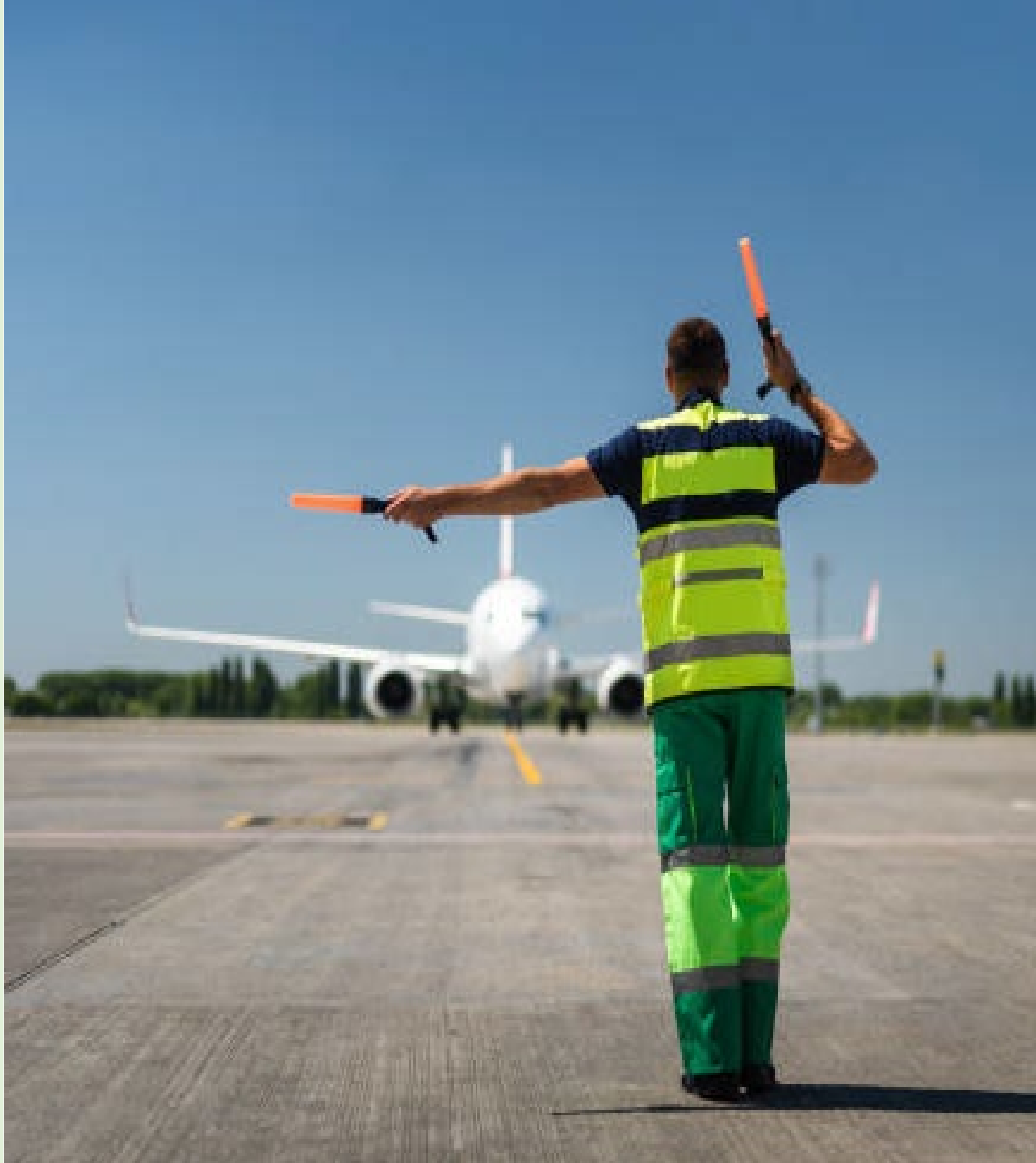
3 key variables:

- Airline
- Time of day
- Day of Week

Agenda

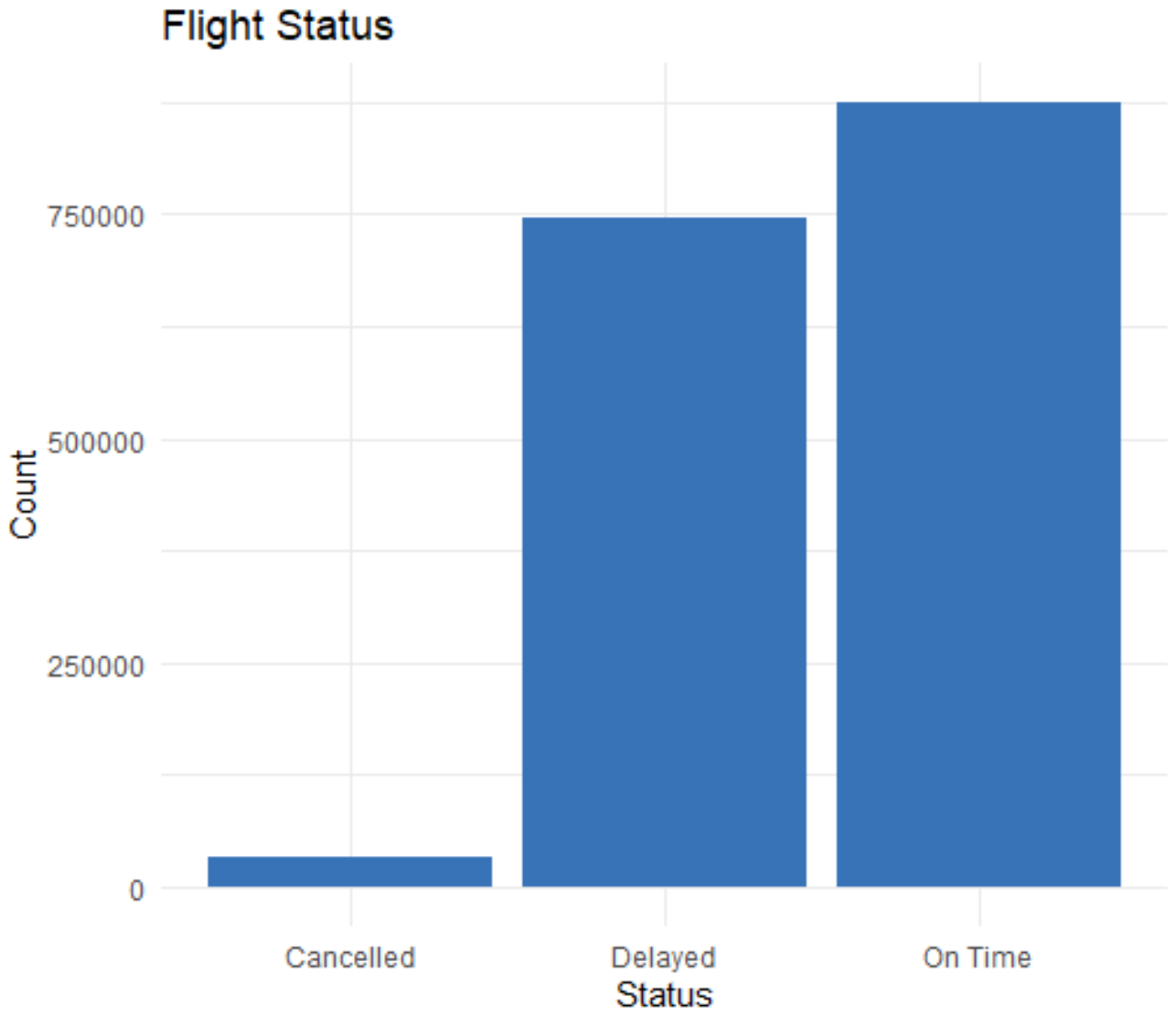
- Background
- Airline vs Delays
- Day of Week vs Delays
- Time of Day vs Delays
- Takeaways

Background



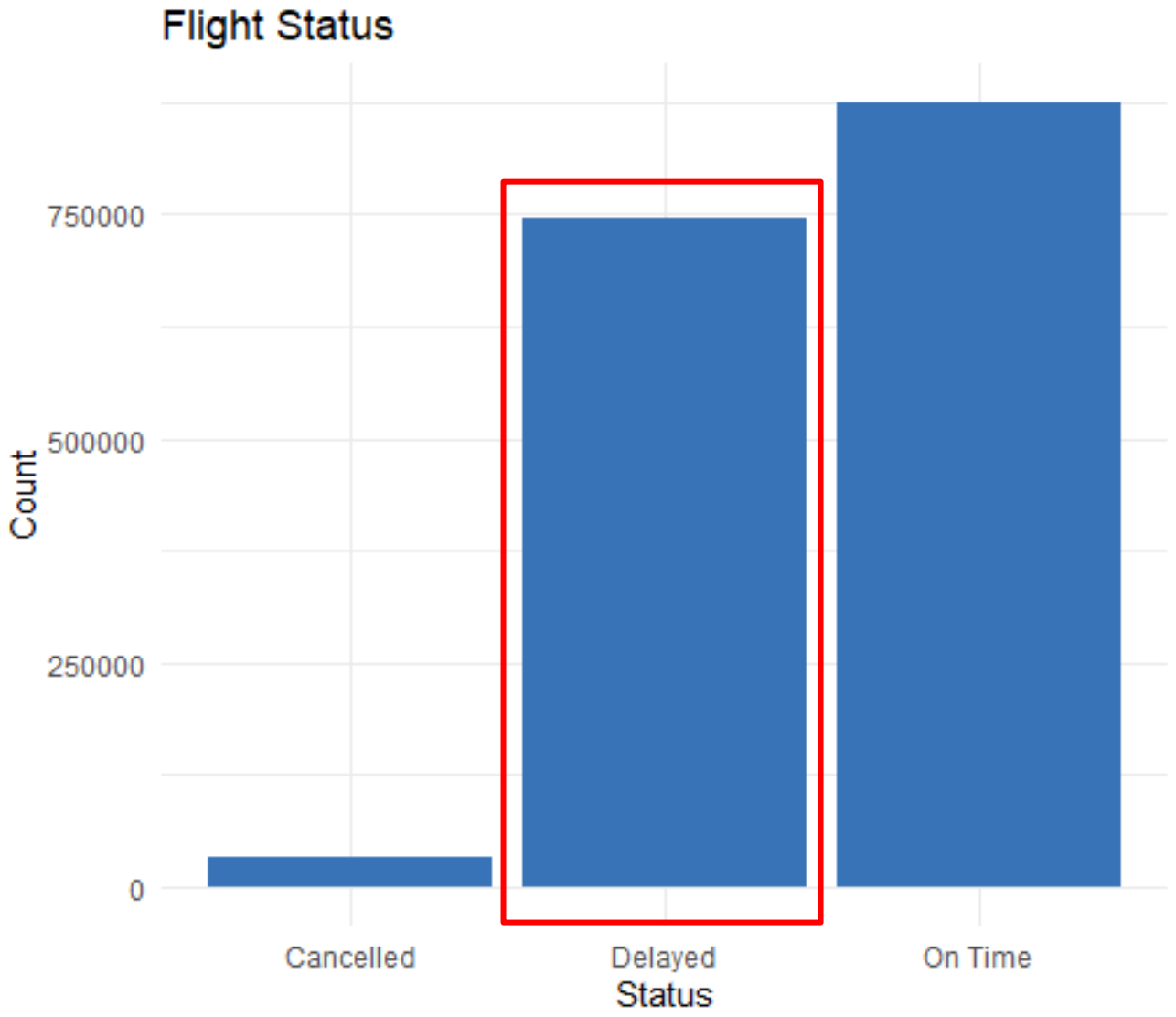
Background

- 1.6 million observations
- 46% delayed



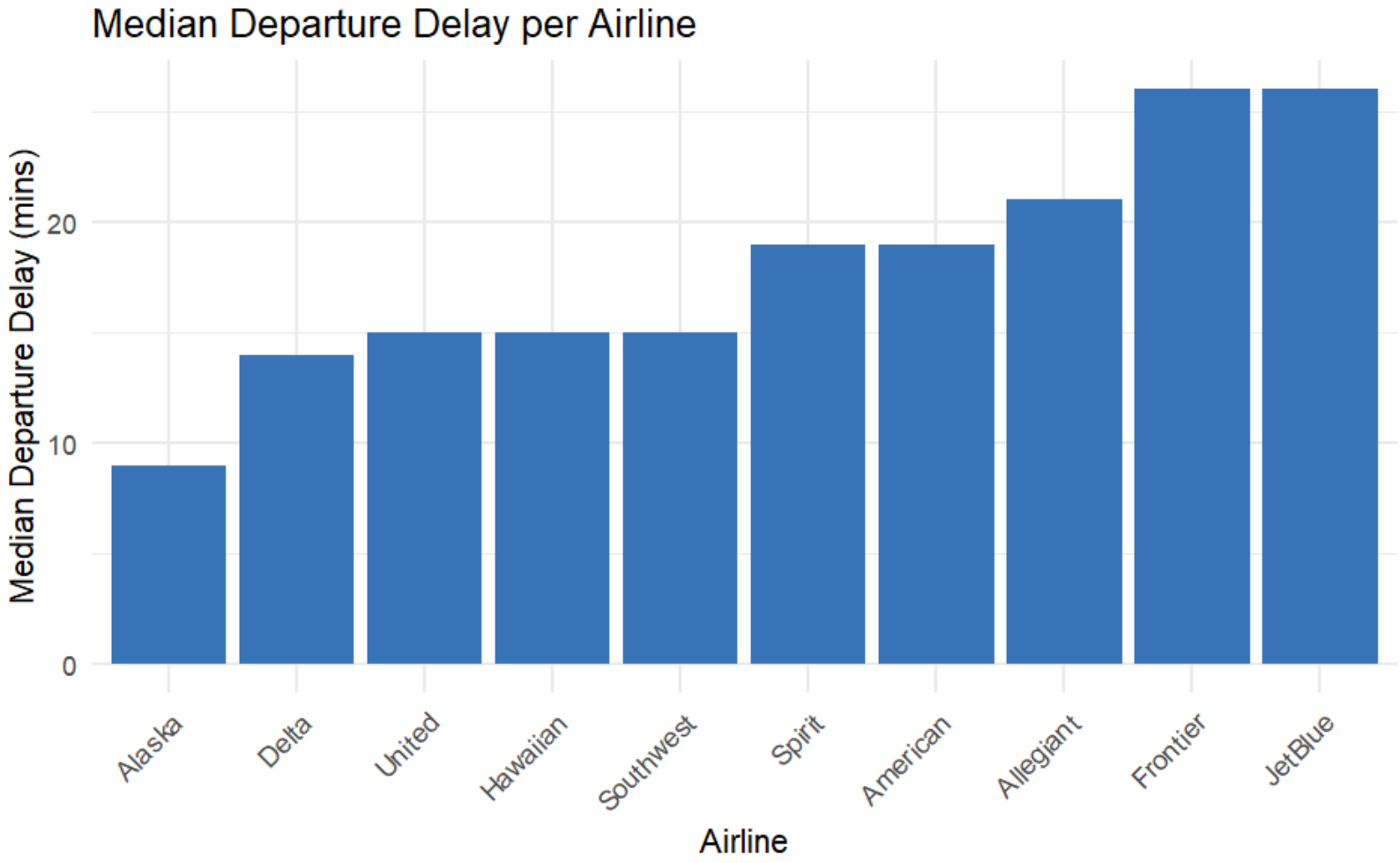
Background

- 1.6 million observations
- 46% delayed
  - Customer impact
  - Airline impact
  - 15/30 "rule"

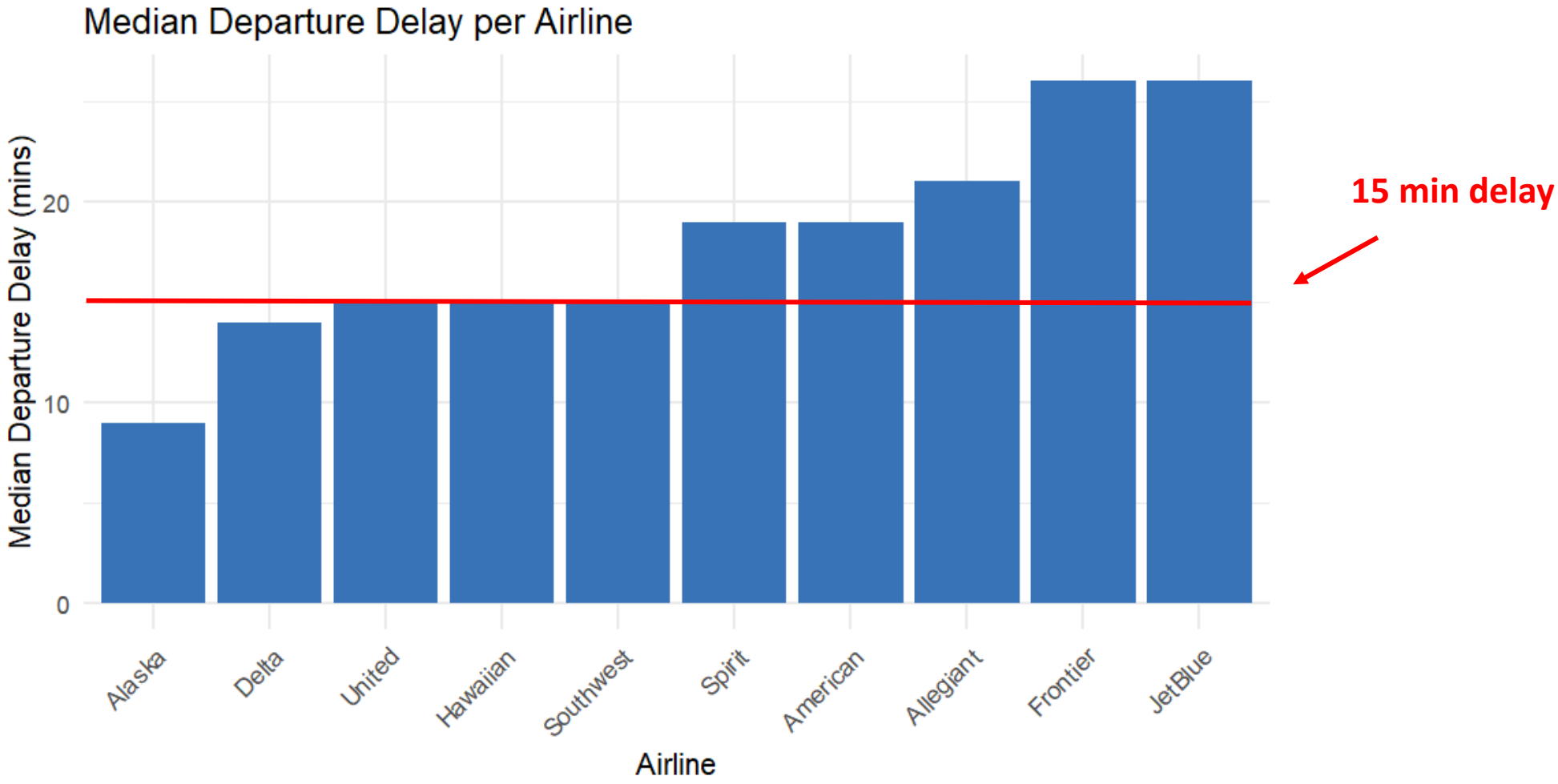


Delays by Airline

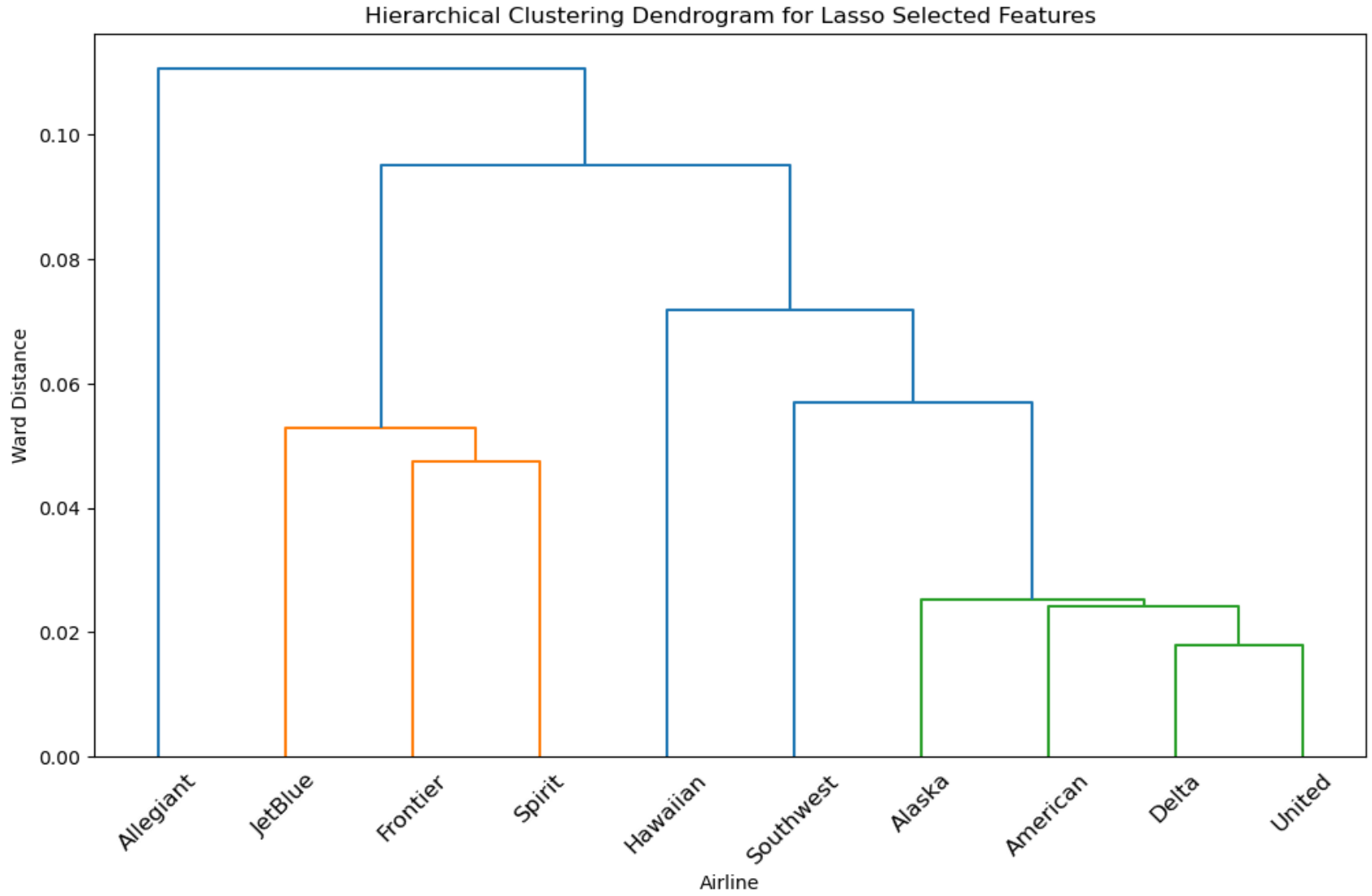
Delays by Airline



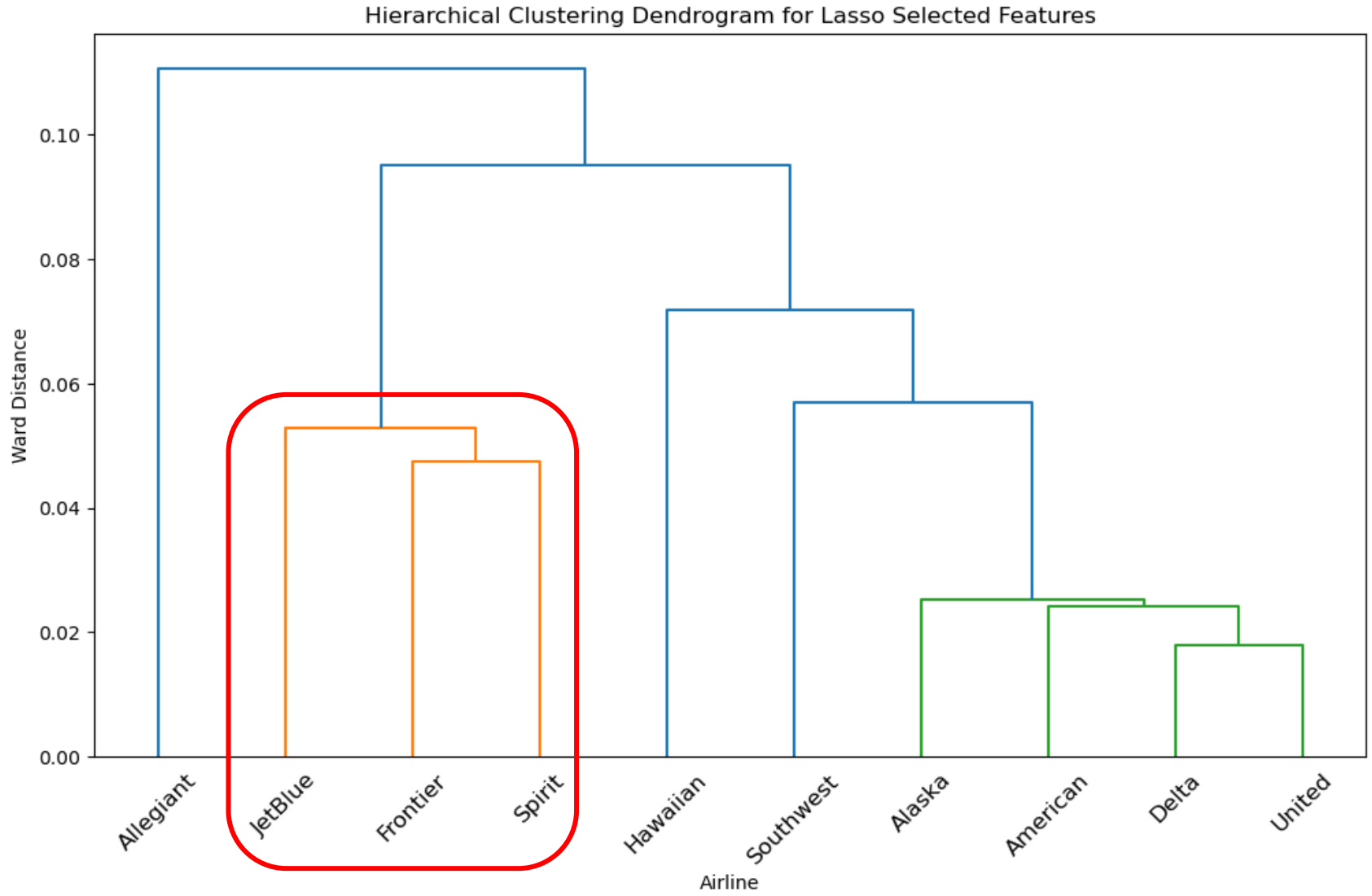
Delays by Airline



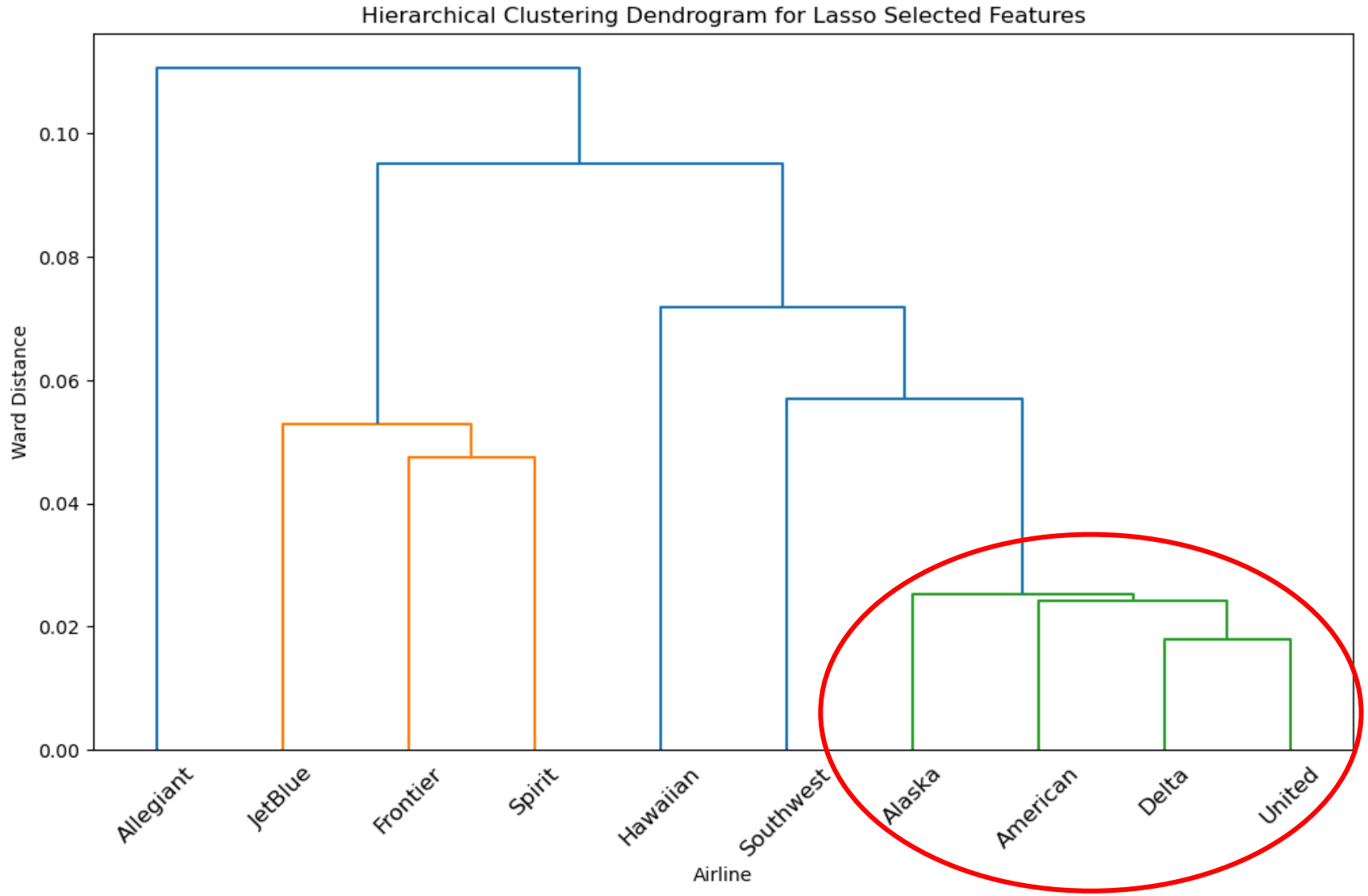
Airline Comparisons



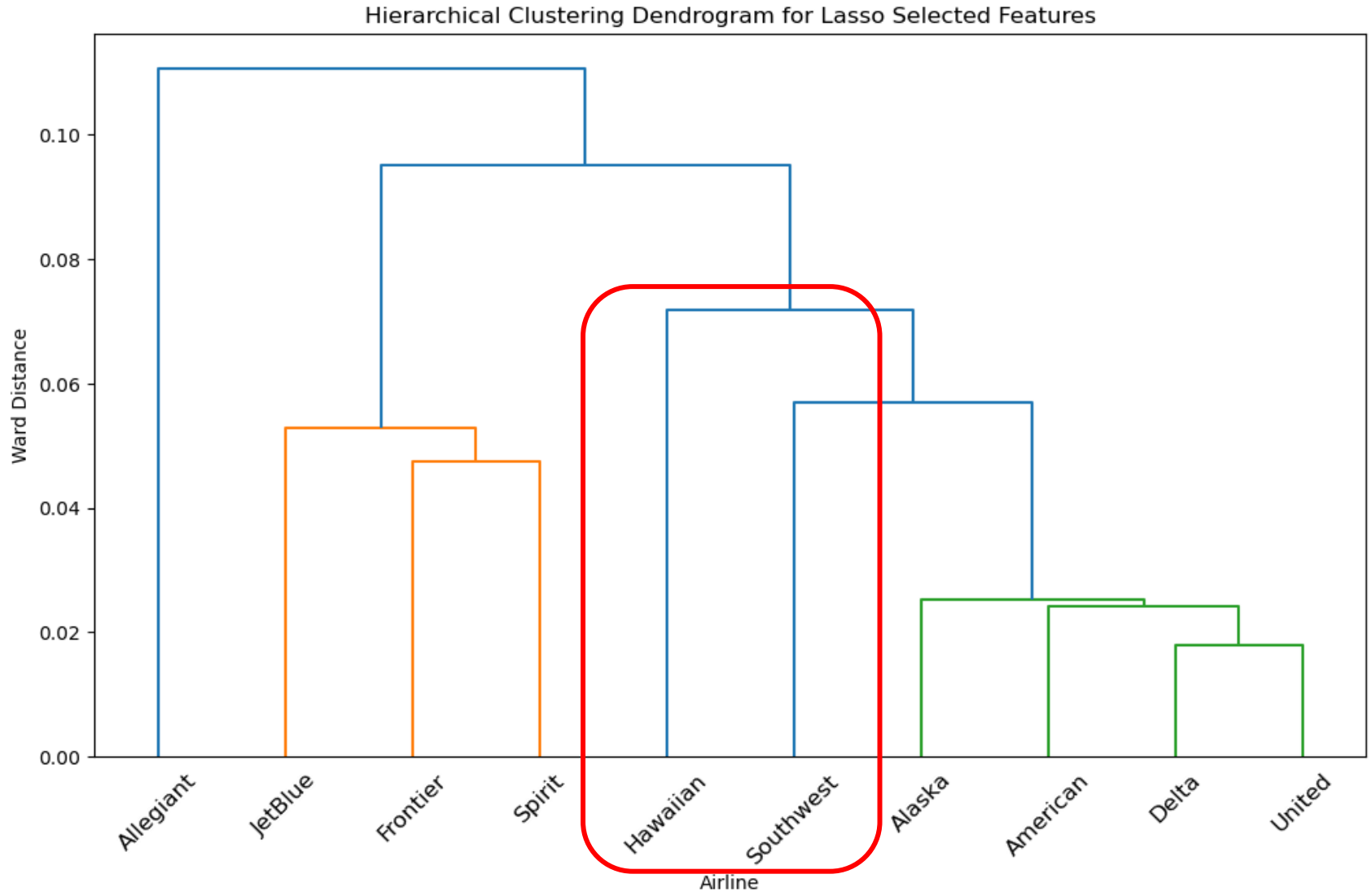
Airline Comparisons



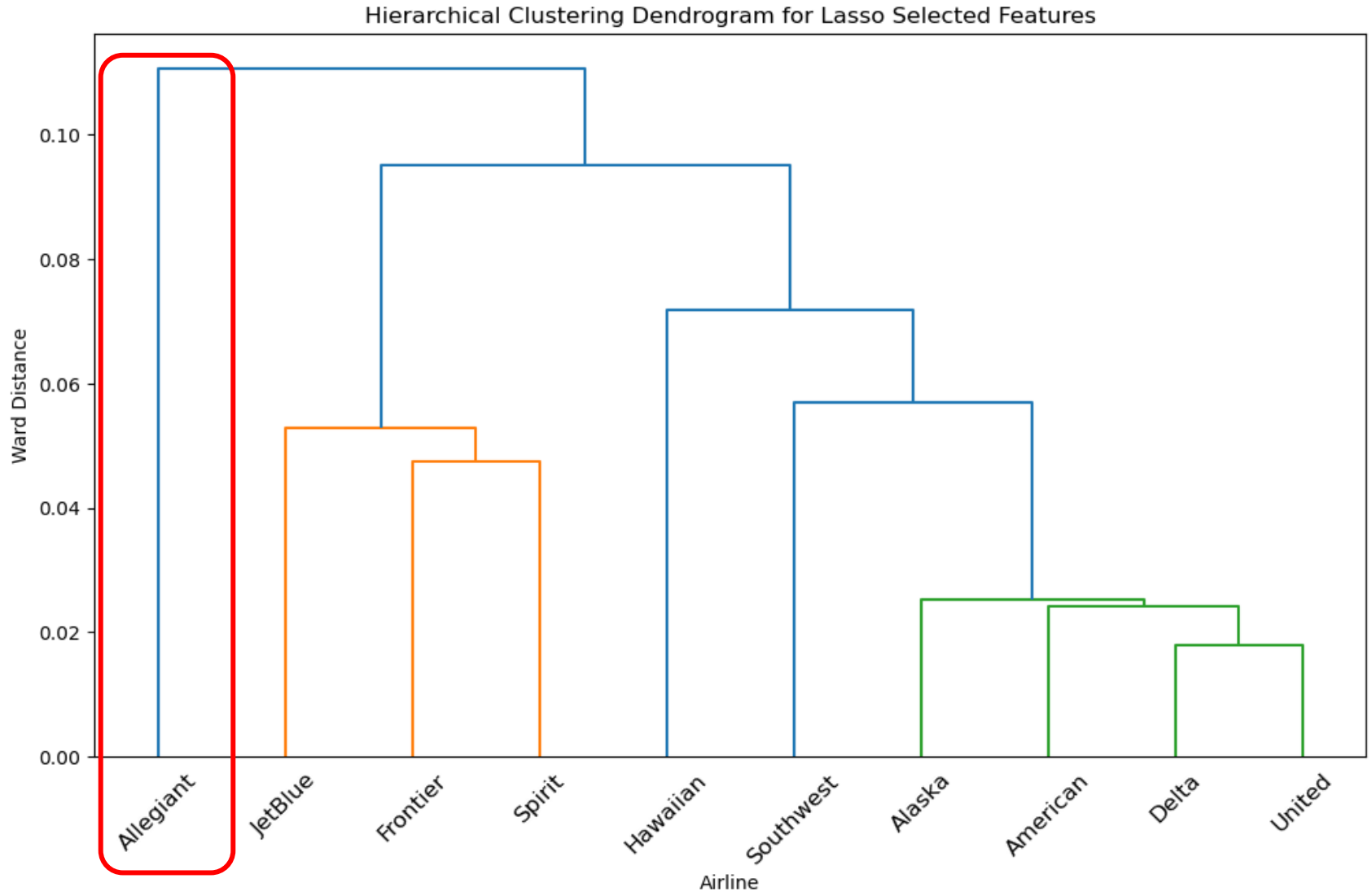
Airline Comparisons



Airline Comparisons



Airline Comparisons

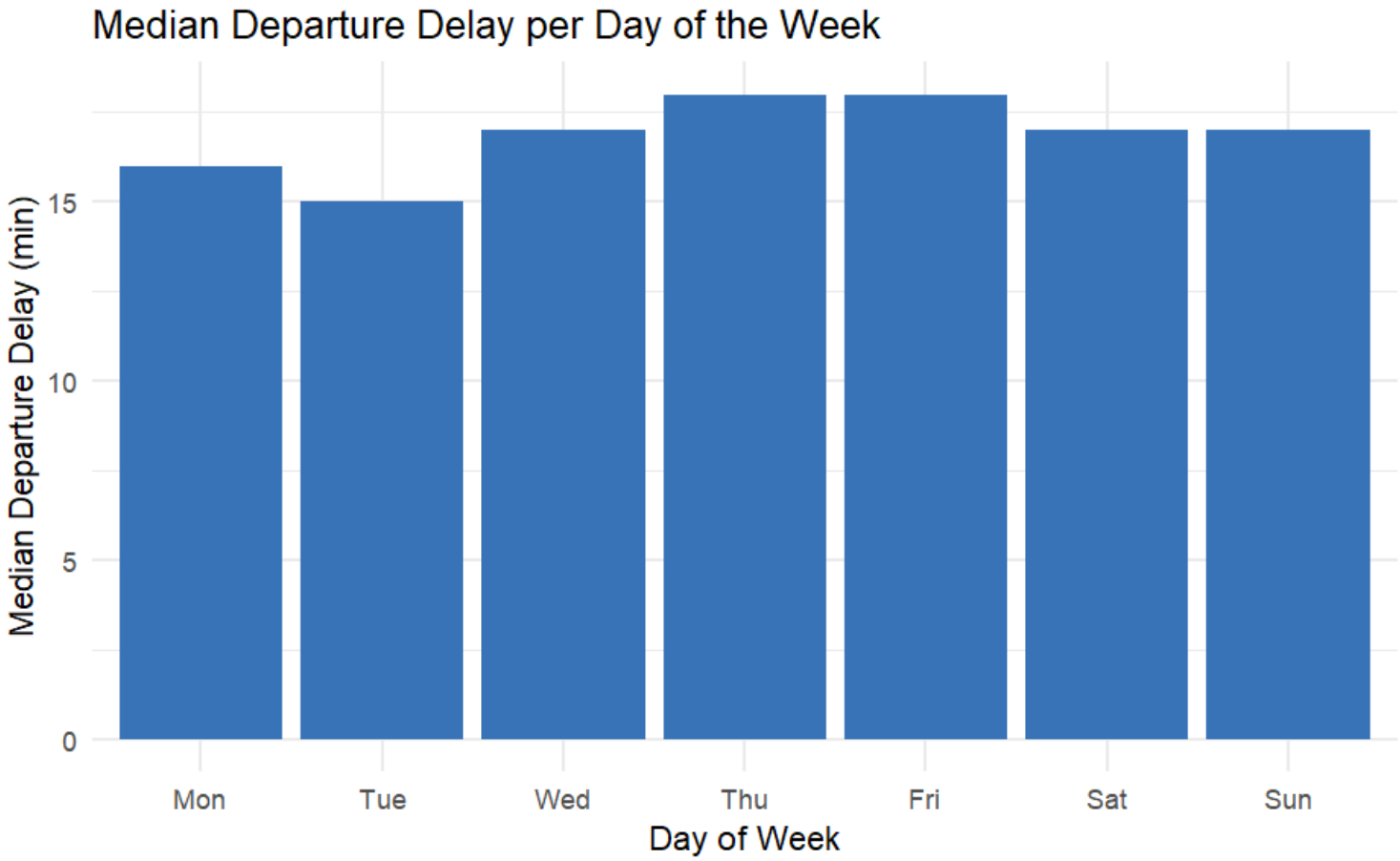


Airline Delay Statistics

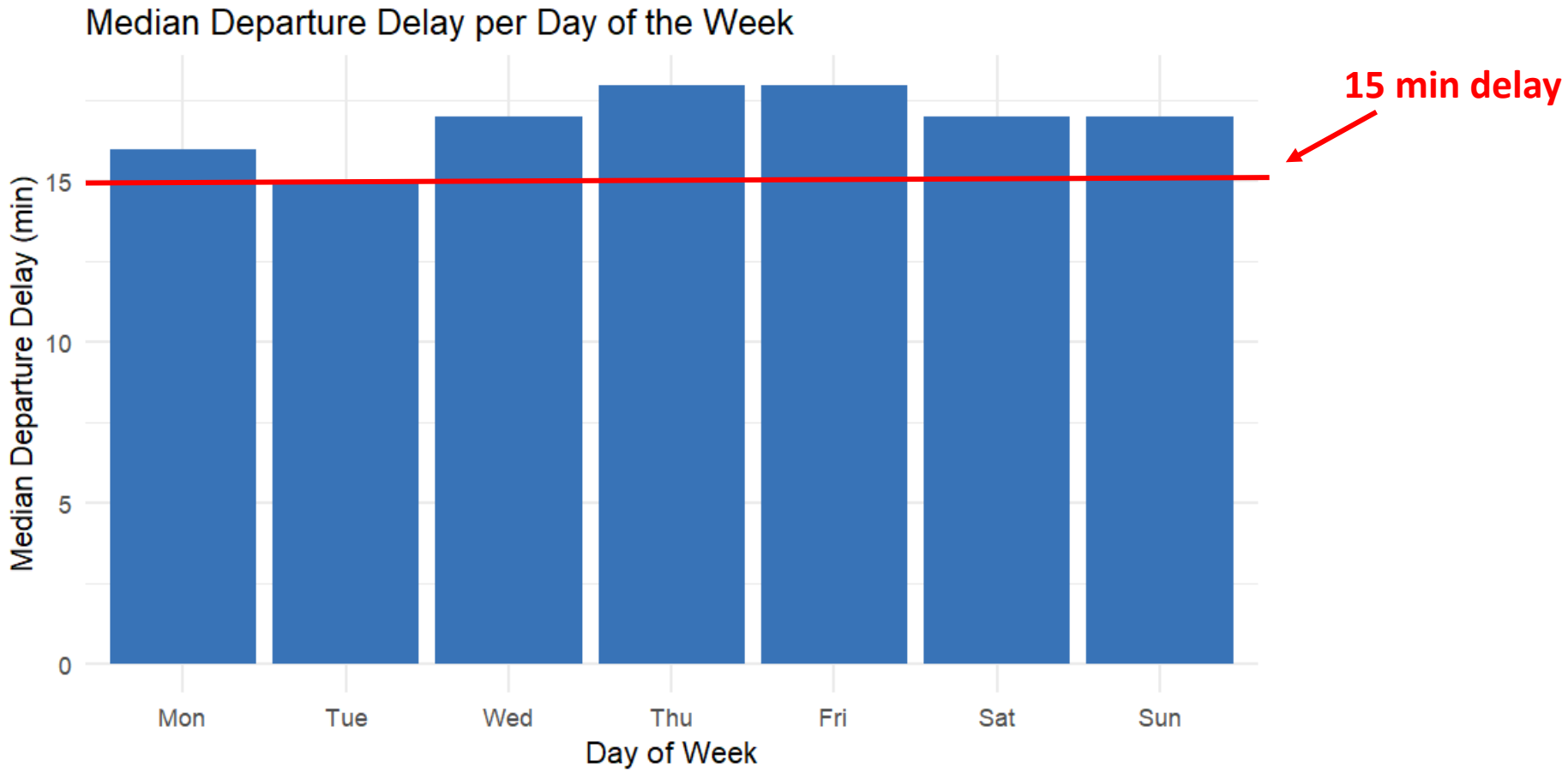
Airline	% Delayed Flights	% Delays > 15 mins	% Delays > 30 mins	Cluster
ALLEGIANT	43%	25%	17%	1
SOUTHWEST	63%	32%	18%	2
HAWAIIAN	44%	16%	9%	
JETBLUE	50%	32%	23%	3
FRONTIER	50%	32%	22%	
SPIRIT	45%	26%	17%	
AMERICAN	43%	24%	16%	4
UNITED	40%	19%	13%	
DELTA	37%	18%	11%	
ALASKA	37%	18%	12%	

Delays by Day of Week

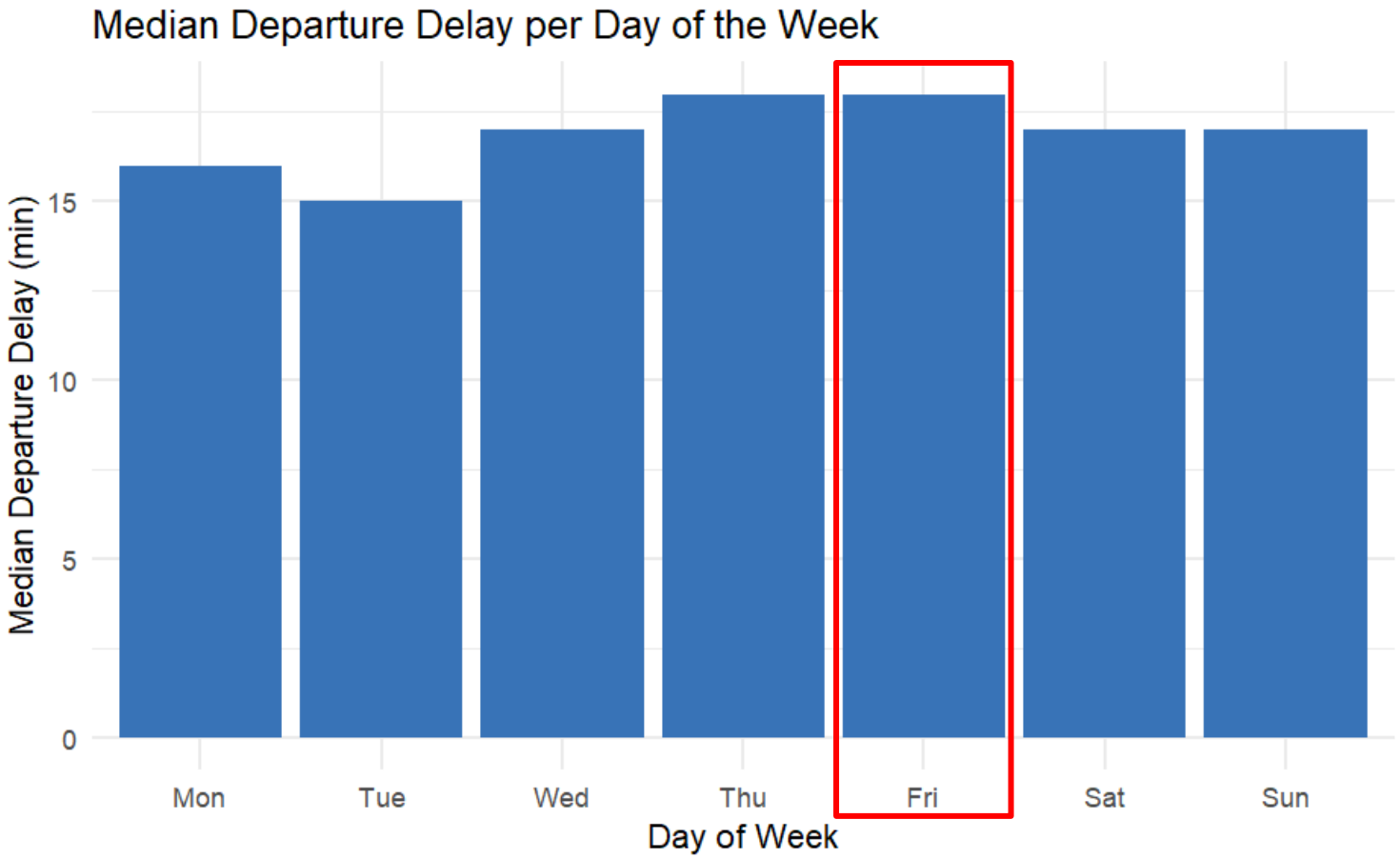
Delays by Day of Week



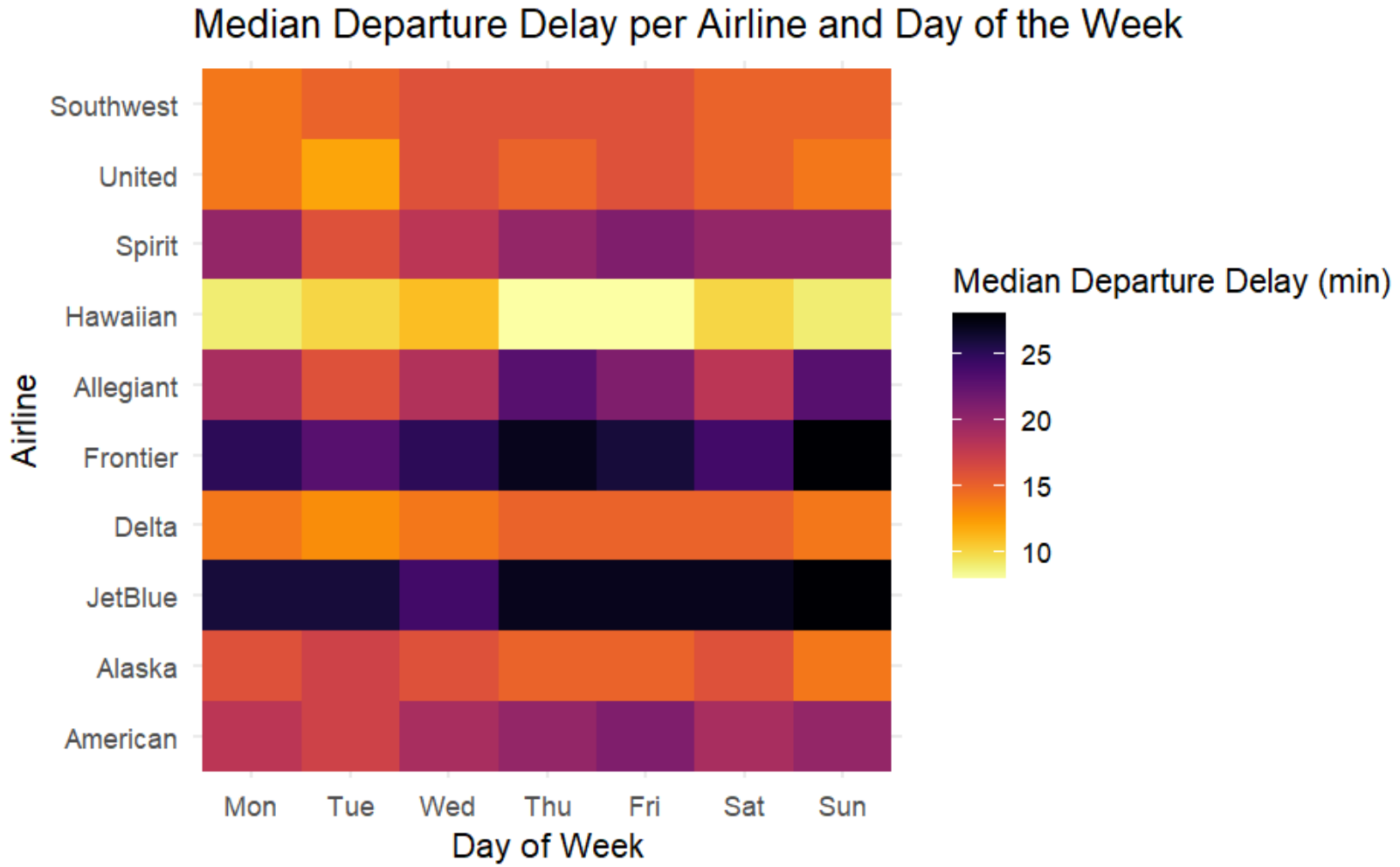
Delays by Day of Week



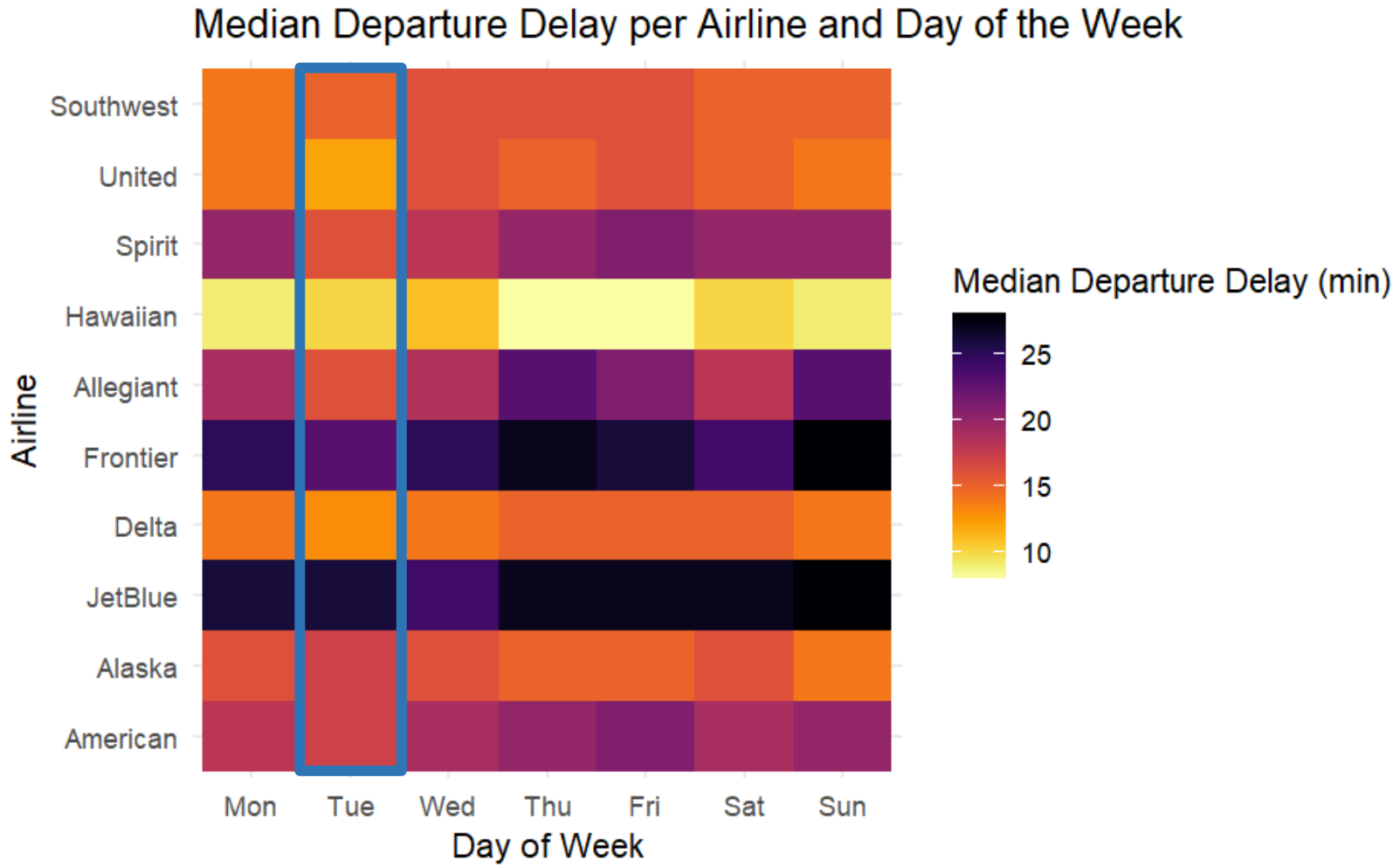
Delays by Day of Week



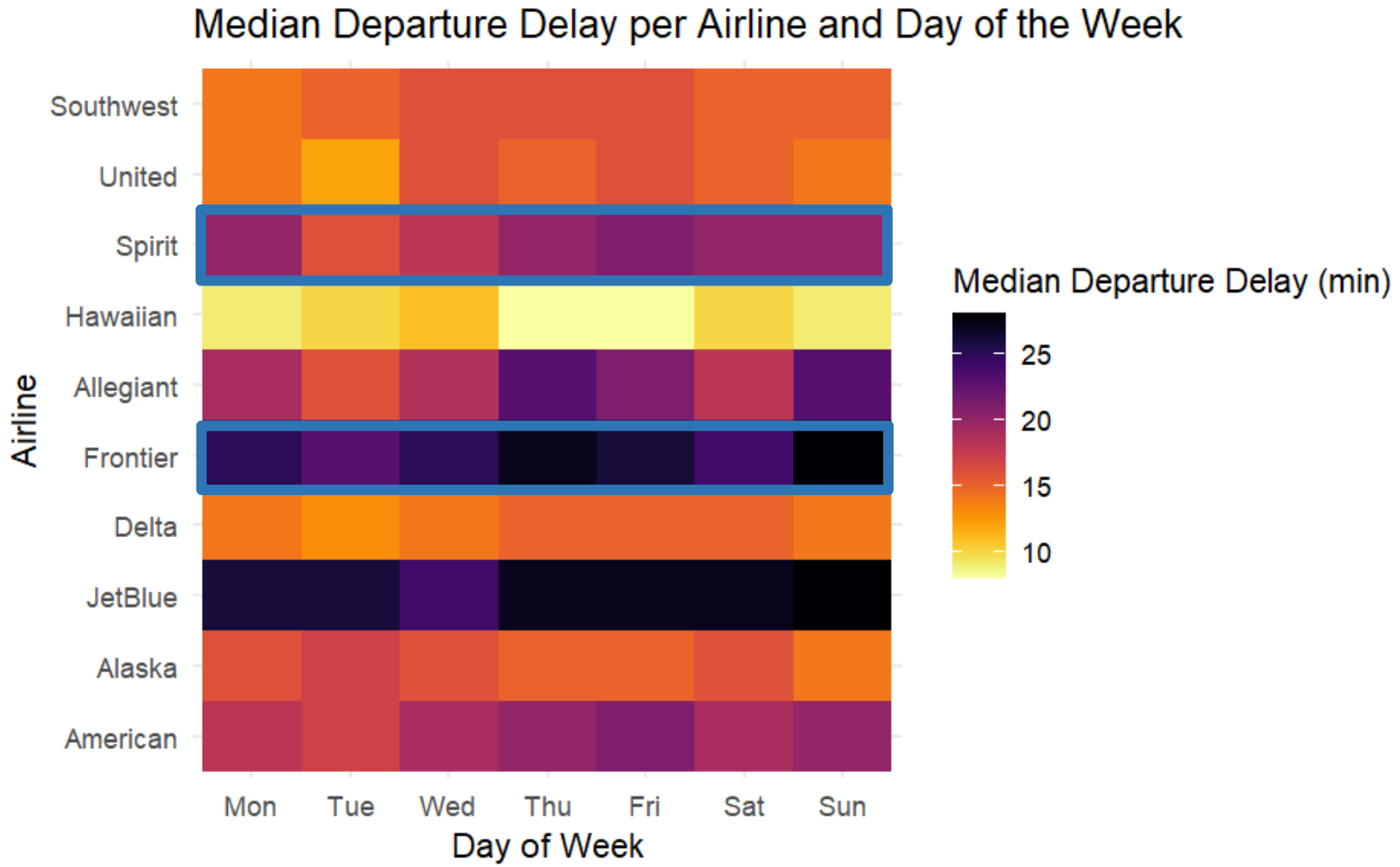
Delays by Airline & Day of Week



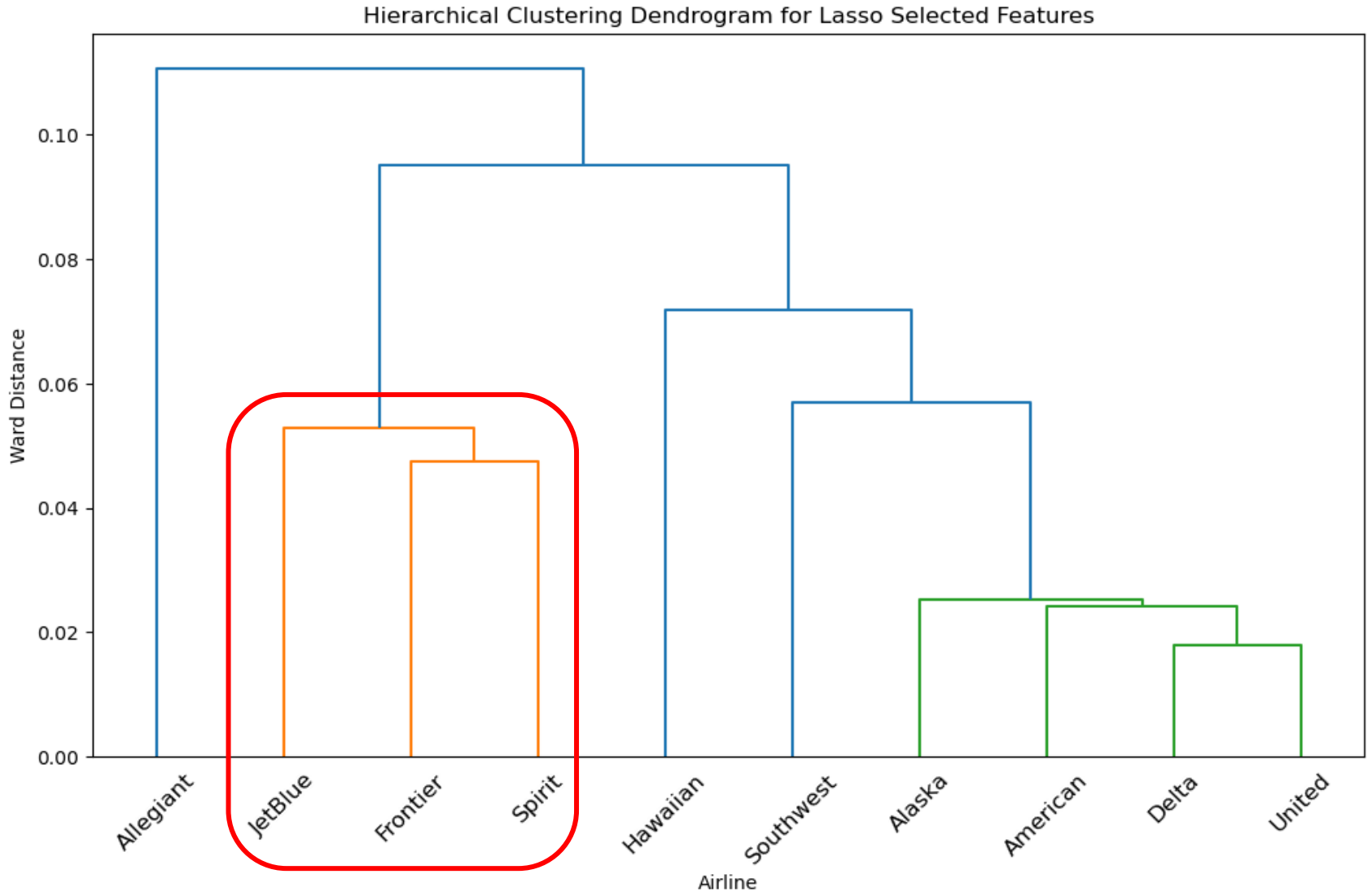
Delays by Airline & Day of Week



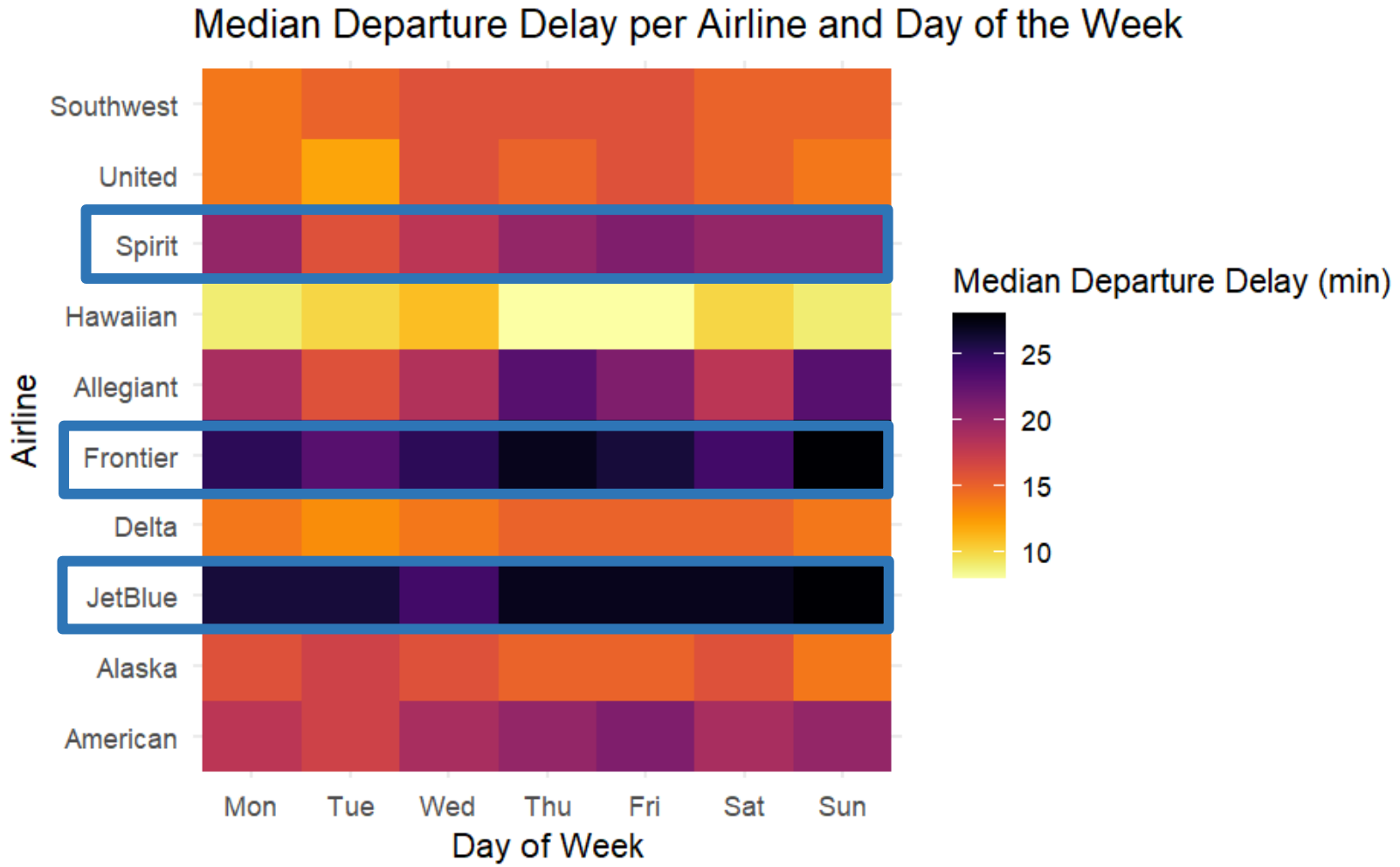
Delays by Airline & Day of Week



Length of Delay by Airline & Day of Week

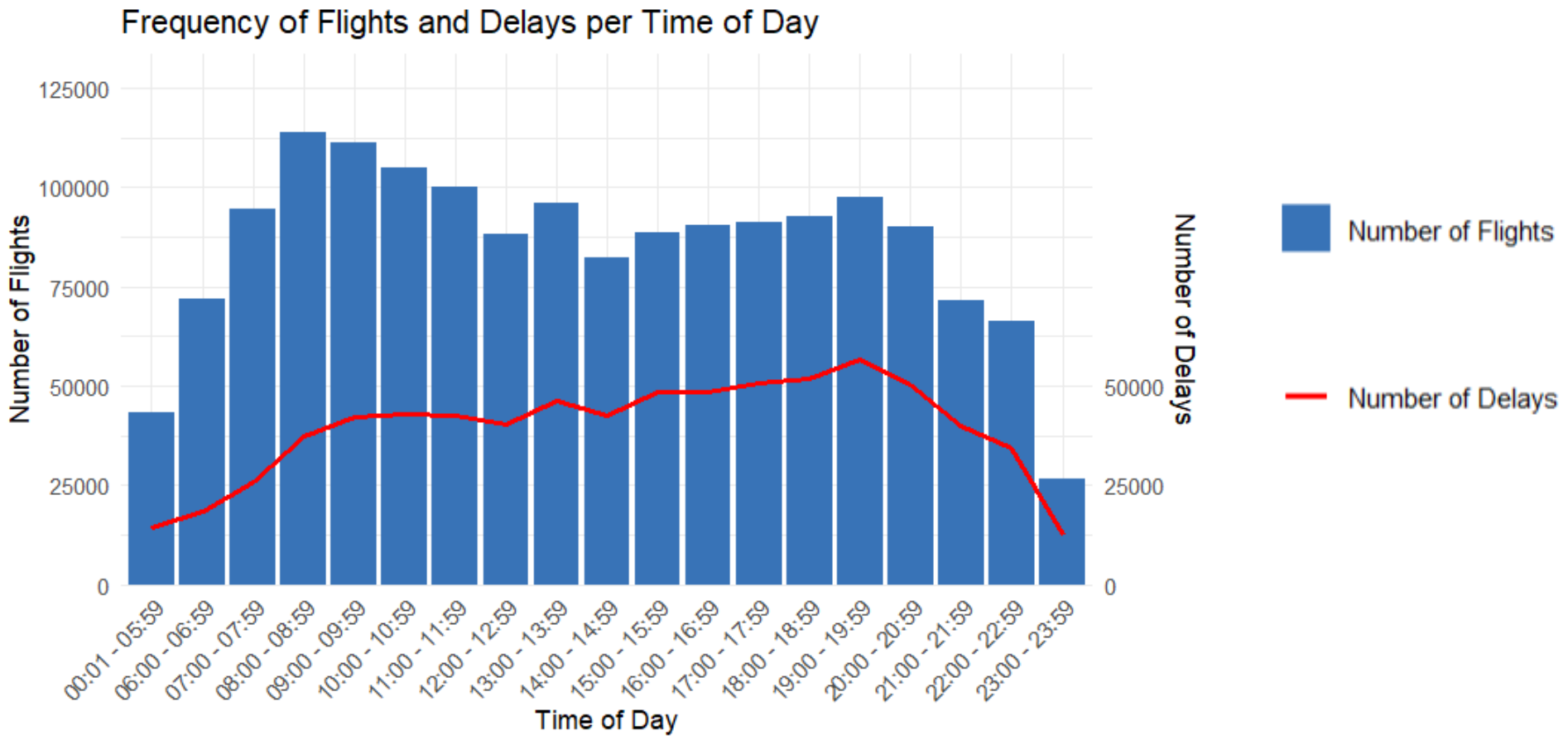


Delays by Airline & Day of Week

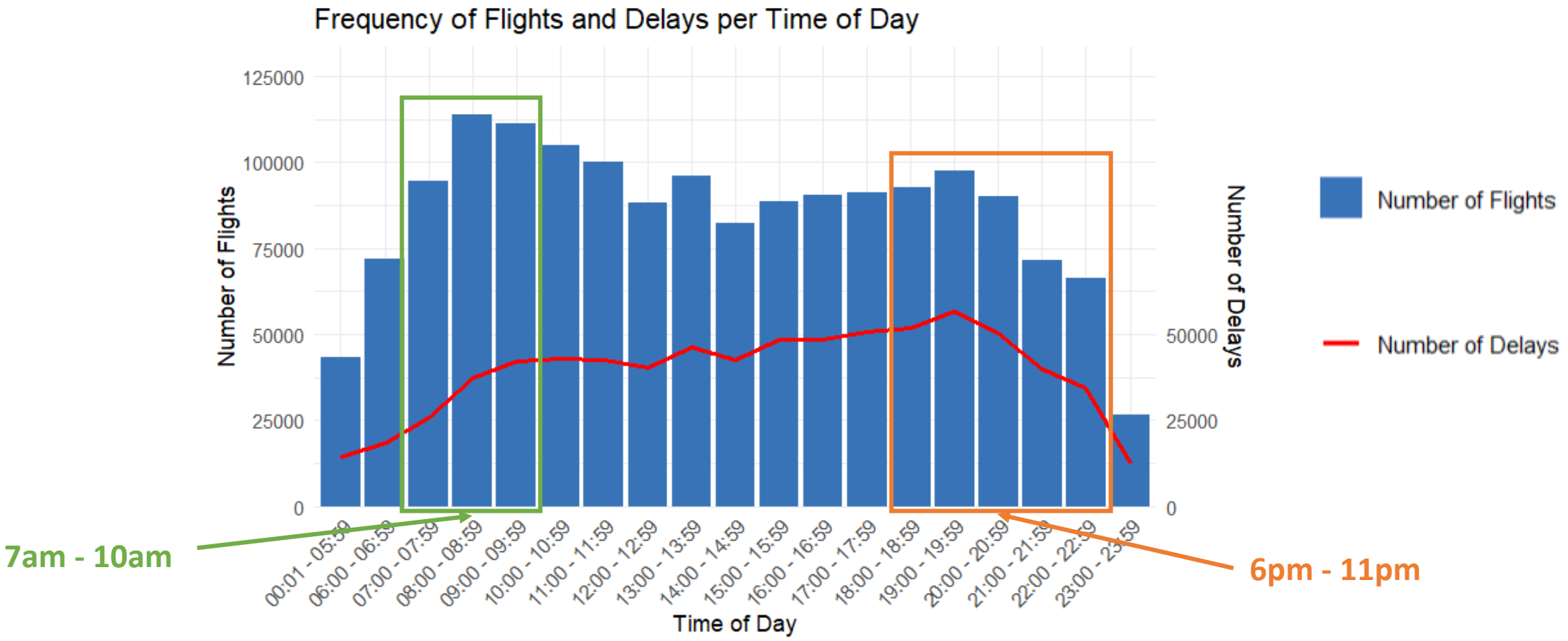


Delays by Time of Day

Peak Flying Hours

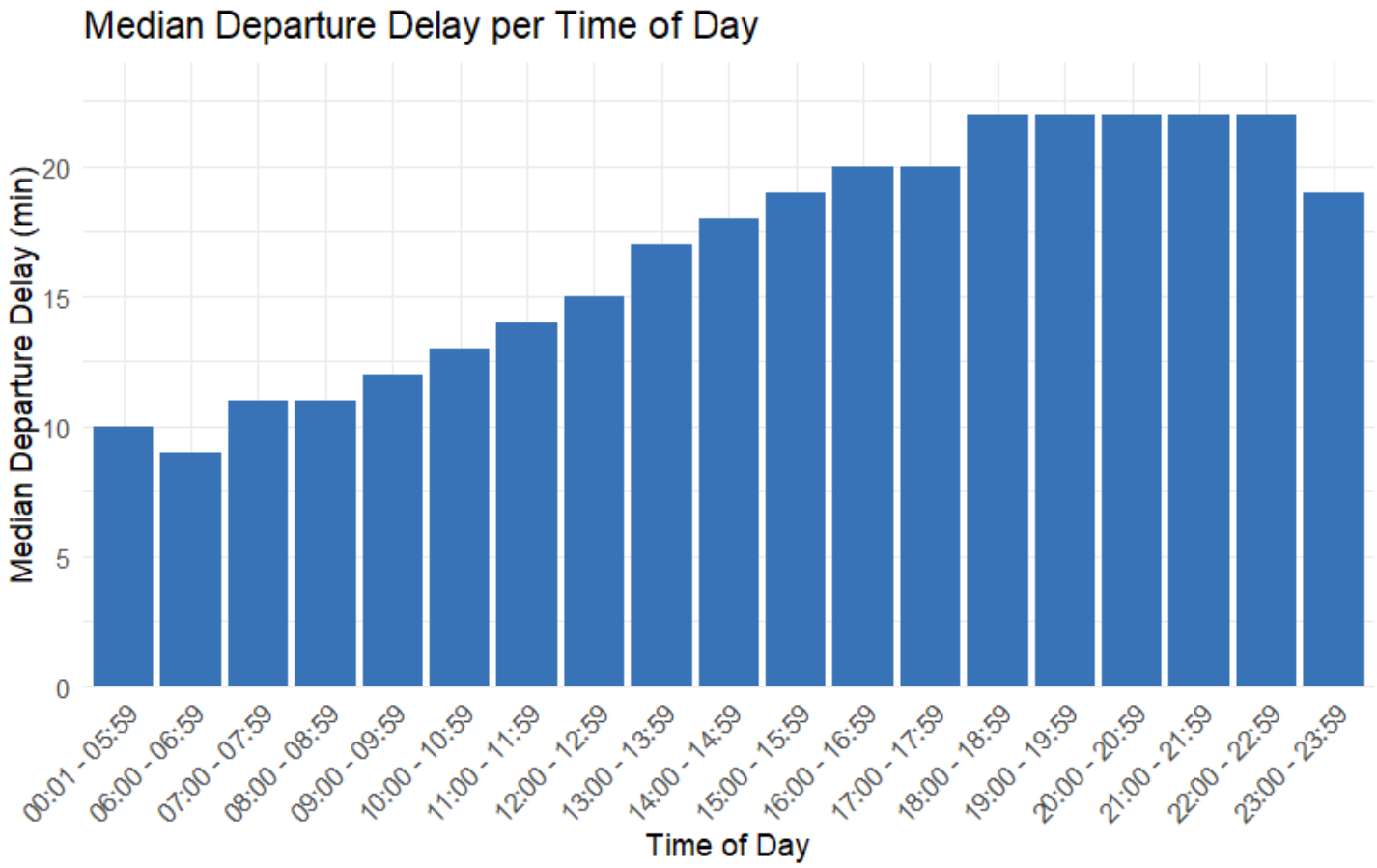


Peak Flying Hours

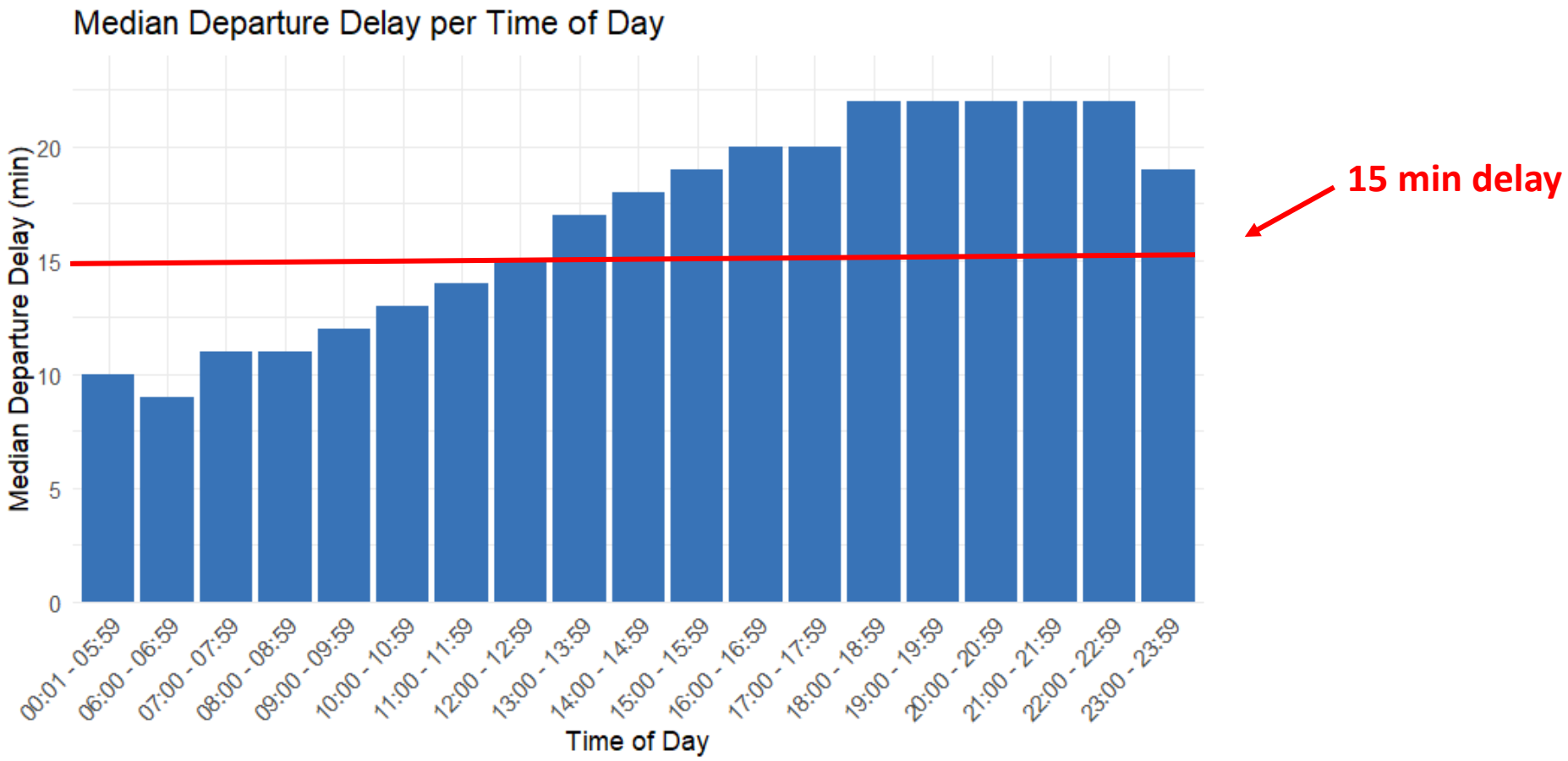


Background → Airline → Day of Week → **Time of Day** → Takeaways

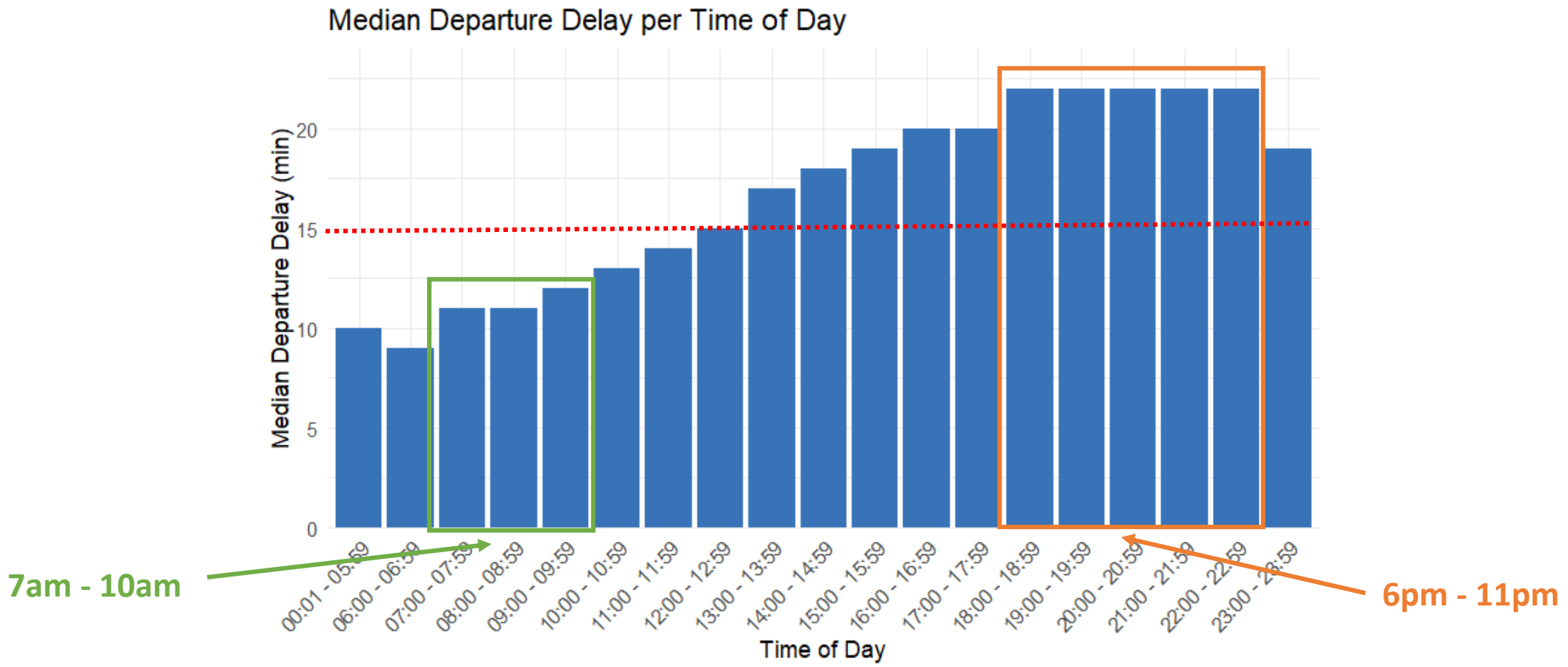
Length of Delay by Time of Day



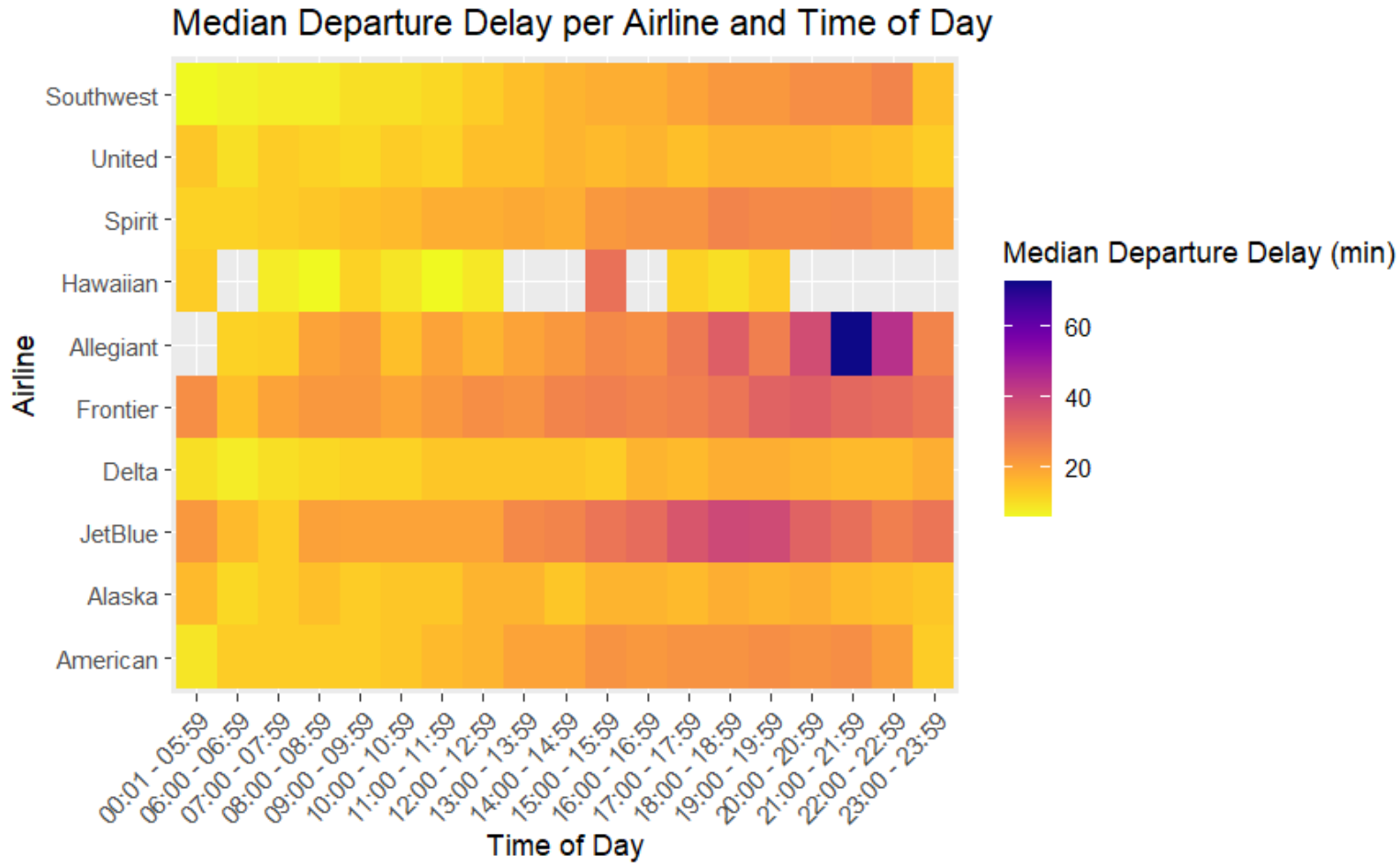
Length of Delay by Time of Day



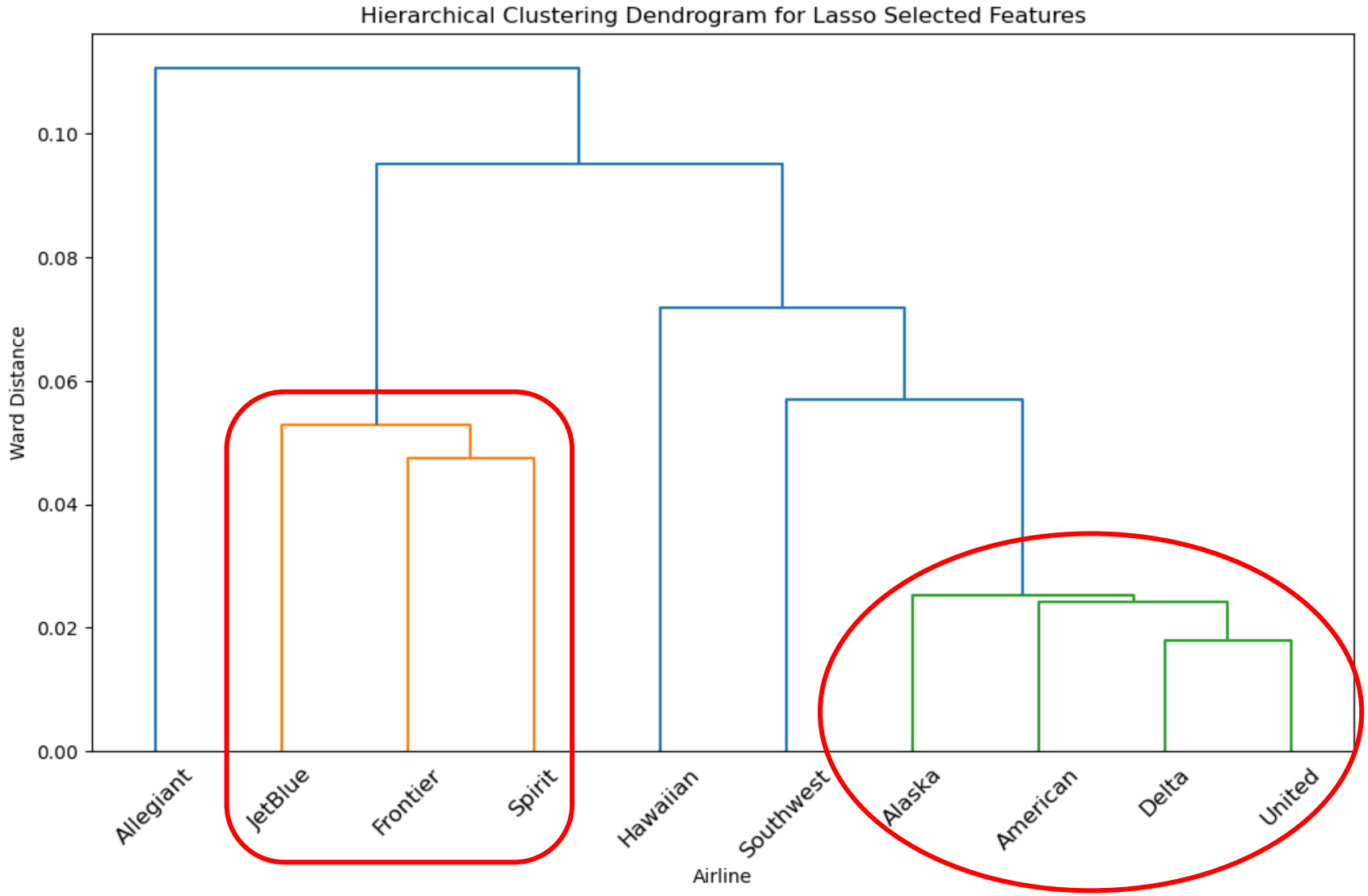
Length of Delay by Time of Day



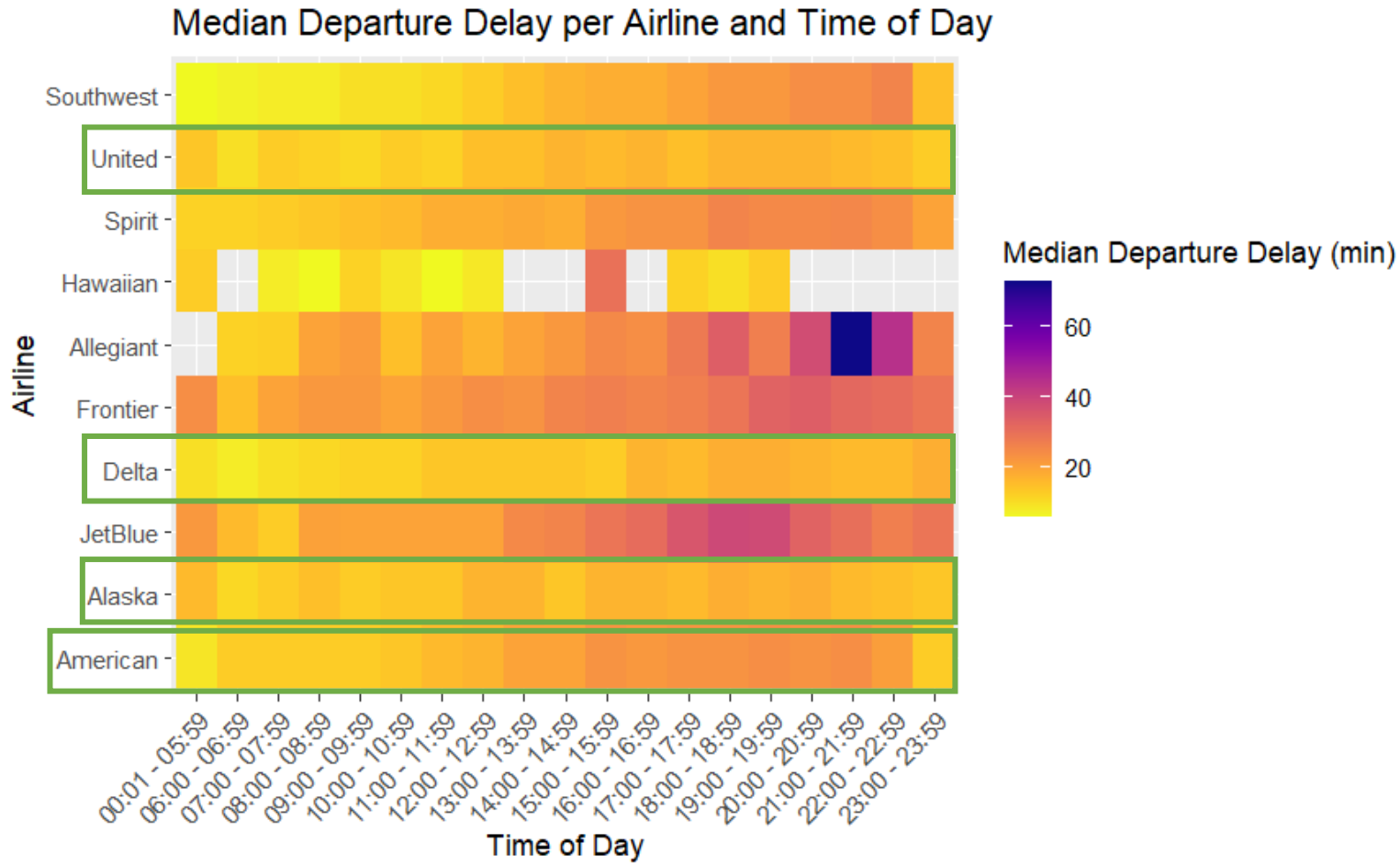
Length of Delay by Airline & Time of Day



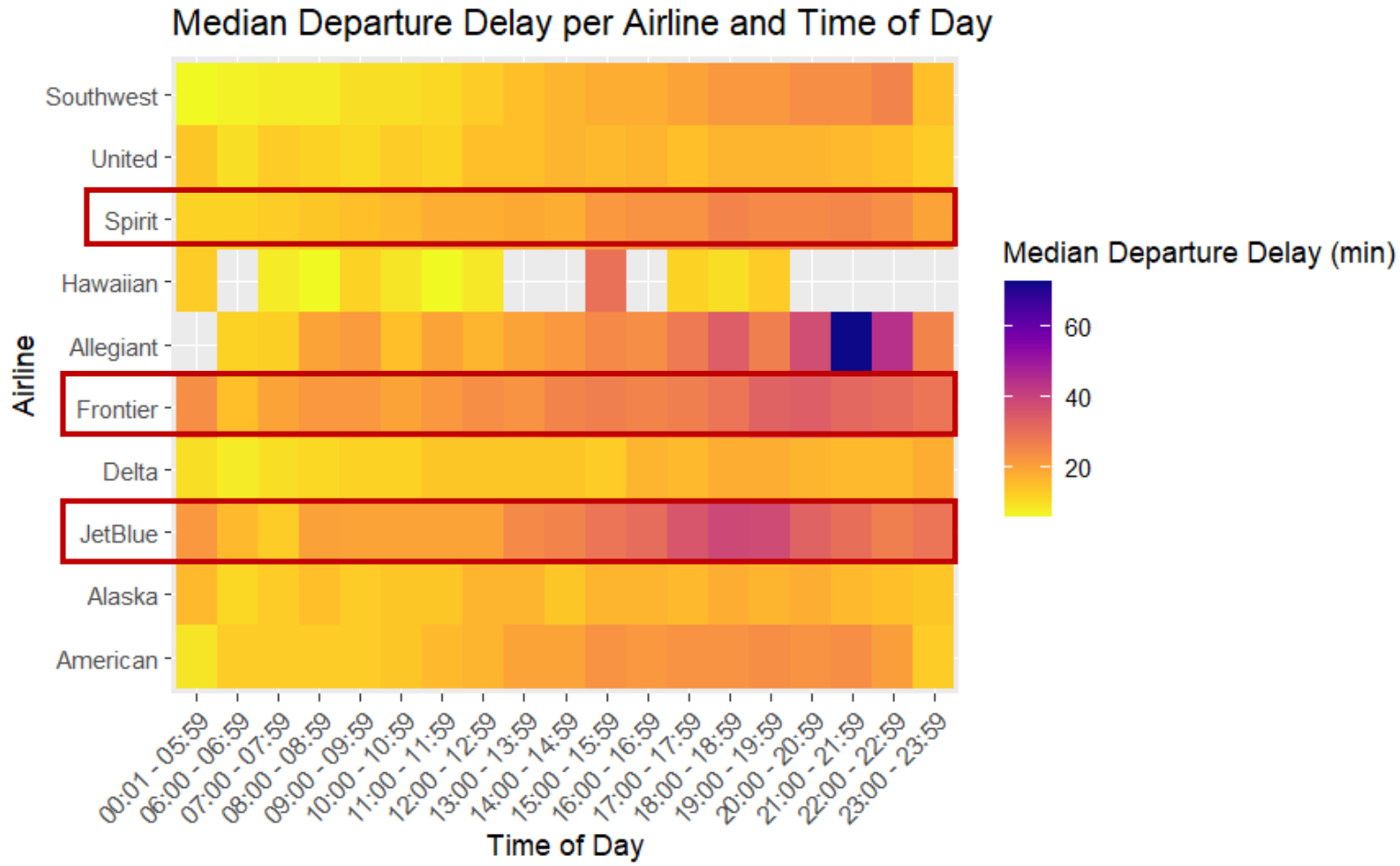
Length of Delay by Airline & Time of Day



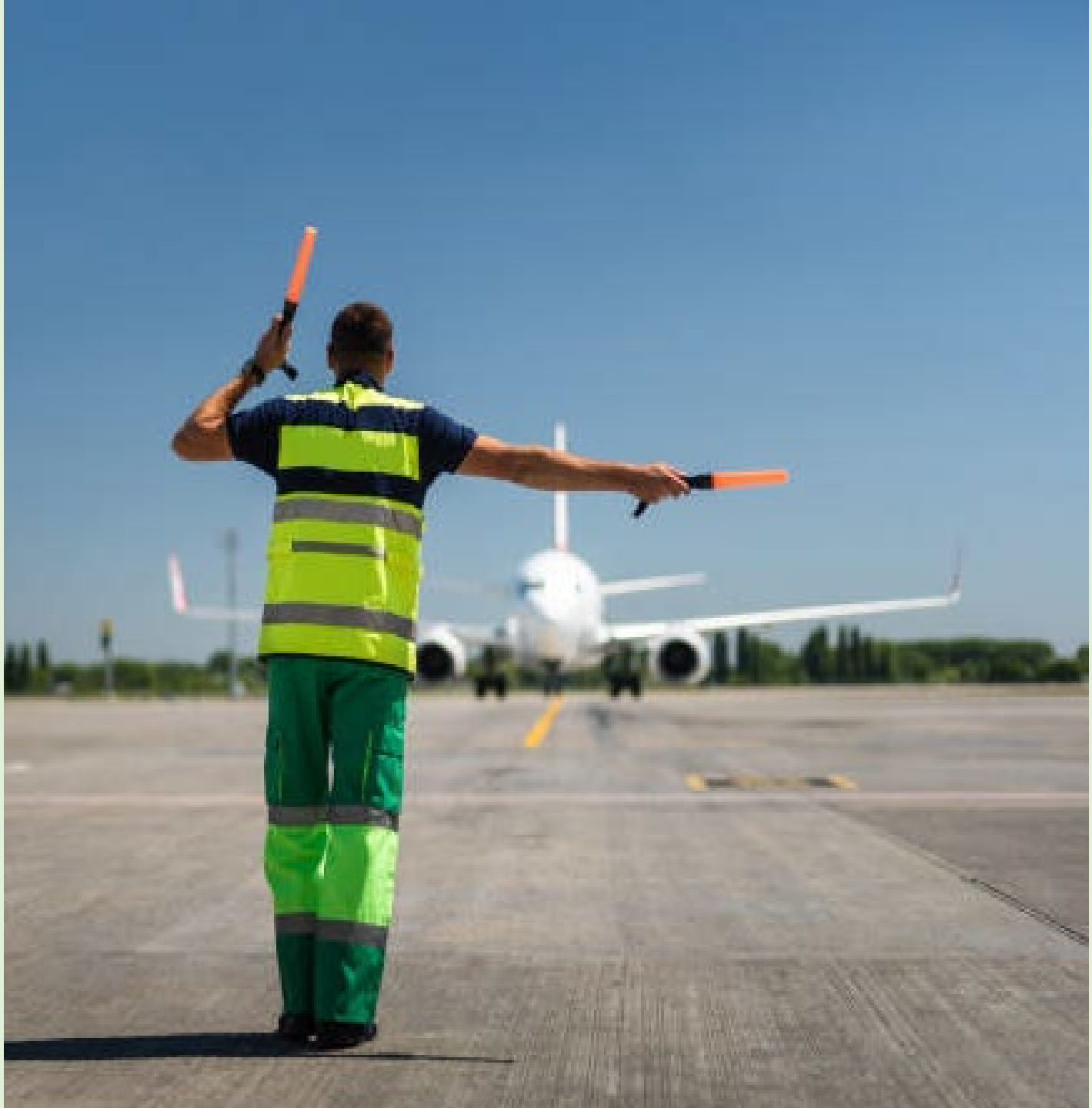
Length of Delay by Airline & Time of Day



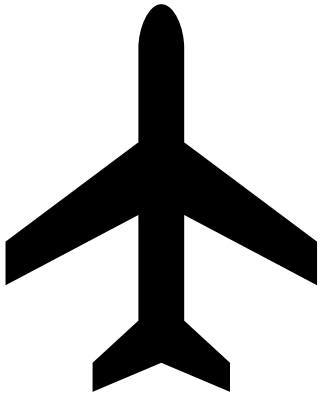
Length of Delay by Airline & Time of Day



Takeaways



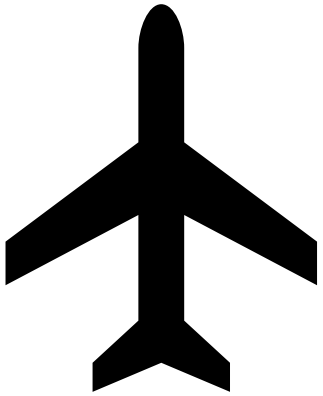
Summary



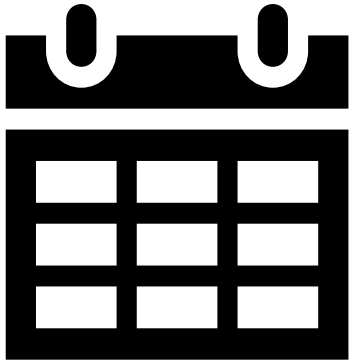
- 2 major performance groups

Background → Airline → Day of Week → Time of Day → **Takeaways**

Summary



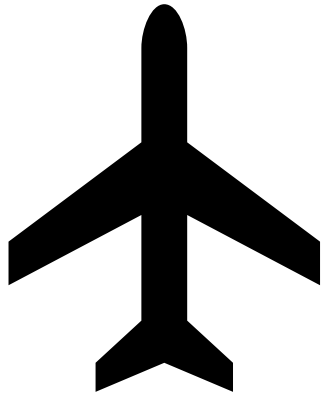
- 2 major performance groups



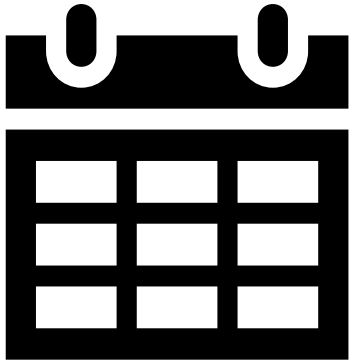
- Tuesday is least delayed
- Friday is most delayed

Background → Airline → Day of Week → Time of Day → **Takeaways**

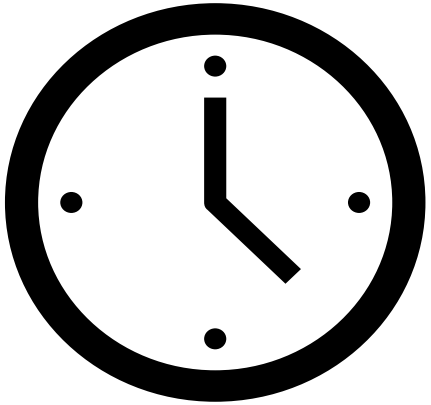
Summary



- 2 major performance groups



- Tuesday is least delayed
- Friday is most delayed



- Delays are amplified by TOD

Recommendations

Re-evaluate

Connection

Time

Guidance

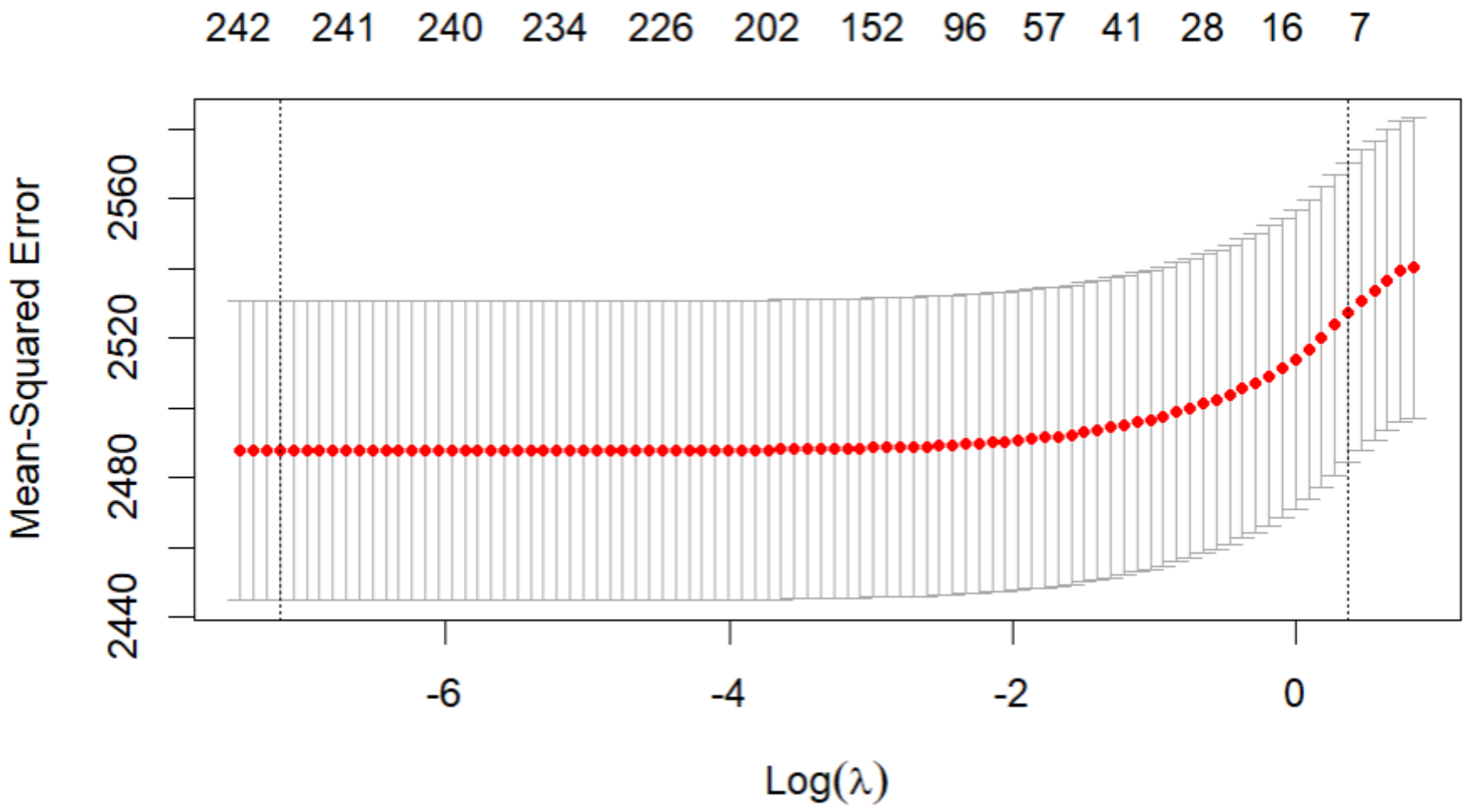
3 dimensions:

- Airline Performance
- Time Blocks
- Day of Week

Now that we have landed
… any questions?

Appendix

Airline - LASSO Lambda



Airline - LASSO Output

```
coef(flight_lasso_cv_drop, s = c( flight_lasso_cv_drop$lambda.halfse,
flight_lasso_cv_drop$lambda.75se, flight_lasso_cv_drop$lambda.90se,
flight_lasso_cv_drop$lambda.1se))
```

```
                              s1           s2           s3         s4
(Intercept)         18.14567540 17.981317851 17.963707262 17.9388925
factor(DayOfWeek)2  -1.35533257 -0.337973941 -0.001269911  .
factor(DayOfWeek)3   .           .            .           .
factor(DayOfWeek)4   .           .            .           .
factor(DayOfWeek)5   .           .            .           .
factor(DayOfWeek)6   .           .            .           .
factor(DayOfWeek)7   .           .            .           .
AirlineAS            .           .            .           .
AirlineB6            5.31297450  3.190702663  1.813972430  0.8978189
AirlineDL           -3.57766374 -2.782115177 -2.343849840 -2.0676816
AirlineF9            3.77610867  2.232675105  1.216915327  0.5116900
AirlineG4            .           .            .           .
AirlineHA            .           .            .           .
AirlineNK            .           .            .           .
AirlineUA           -0.32181841  .            .           .
AirlineWN            .           .            .           .
OriginCLT            .           .            .           .
OriginDEN            .           .            .           .
OriginDFW            0.03407232  .            .           .
OriginLAS            .           .            .           .
OriginLAX            .           .            .           .
```

Airline - LASSO Output

```
coef(flight_lasso_cv_drop, s = c( flight_lasso_cv_drop$lambda.halfse,
flight_lasso_cv_drop$lambda.75se, flight_lasso_cv_drop$lambda.90se,
flight_lasso_cv_drop$lambda.1se))
```

```
DepTimeBlk0600-0659 -6.95326089 -5.119559367 -4.046502041 -3.2807356
DepTimeBlk0700-0759 -5.65374482 -4.214500801 -3.365242590 -2.7488337
DepTimeBlk0800-0859 -5.07647744 -3.850781395 -3.134220597 -2.6060379
DepTimeBlk0900-0959 -3.30280947 -2.047302935 -1.306898377 -0.7629330
DepTimeBlk1000-1059 -2.11693488 -0.769774359 -0.052925502  .
DepTimeBlk1100-1159 -1.06036628 -0.002708109  .            .
DepTimeBlk1200-1259  .           .            .            .
DepTimeBlk1300-1359  .           .            .            .
DepTimeBlk1400-1459  .           .            .            .
DepTimeBlk1500-1559  .           .            .            .
DepTimeBlk1600-1659  0.39371586  .            .            .
DepTimeBlk1700-1759  2.84237352  1.296982631  0.359816821  .
DepTimeBlk1800-1859  3.12973585  1.634836517  0.716685749  0.1489059
DepTimeBlk1900-1959  4.55895228  3.073603337  2.170413471  1.6136670
DepTimeBlk2000-2059  3.23575585  1.737461003  0.810474088  0.2368390
DepTimeBlk2100-2159  2.50024030  0.836913705  0.037596804  .
DepTimeBlk2200-2259  2.60832623  0.819674065  0.029153727  .
DepTimeBlk2300-2359  .           .            .            .
Distance             .           .            .            .
```

Airline - Model R^2

```
r_squared_test <- 1 - sum((test_y_drop - test_pred)^2) / sum((test_y_drop -
mean(test_y_drop))^2)

```

```
[1] 0.006677327
```

Airline - ANOVA & Post Hoc of Delay

```
                    Df    Sum Sq Mean Sq F value Pr(>F)
factor(Airline)      9 2.265e+07 2516749   533.4 <2e-16 ***
Residuals       447948 2.113e+09    4718
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

        Kruskal-Wallis rank sum test

data:  DepDelayMinutes by factor(Airline)
Kruskal-Wallis chi-squared = 6141.2, df = 9, p-value < 2.2e-16


        Pairwise comparisons using Games-Howell test

data: DepDelayMinutes by factor(Airline)

   AA       AS       B6       DL       F9       G4       HA       NK       UA
AS 2.4e-11  -        -        -        -        -        -        -        -
B6 1.0e-08  3.2e-08  -        -        -        -        -        -        -
DL < 2e-16  1.0000   1.4e-08  -        -        -        -        -        -
F9 6.9e-14  6.0e-09  < 2e-16  < 2e-16  -        -        -        -        -
G4 0.9999   7.2e-07  2.3e-05  2.4e-07  0.5015   -        -        -        -
HA 0.9901   0.8967   0.0055   0.9133   0.3630   0.9728   -        -        -
NK 3.8e-06  < 2e-16  2.0e-08  < 2e-16  < 2e-16  0.6307   1.0000   -        -
UA < 2e-16  2.2e-10  2.7e-08  < 2e-16  < 2e-16  0.0541   1.0000   0.0011   -
WN < 2e-16  7.7e-13  < 2e-16  < 2e-16  < 2e-16  < 2e-16  0.0964   < 2e-16  < 2e-16
```

Airline - Dendrogram Input

	DepDelayMinutesMedian																	
DayOfWeek	1																	
DepHour	0	1	2	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Airline																		
Alaska	0.005349	0.000000	0.000000	0.001337	0.003009	0.005349	0.004514	0.004346	0.004179	0.004346	0.007355	0.004681	0.005349	0.005684	0.007355	0.005015	0.006352	0.0
Allegiant	0.000000	0.000000	0.000000	0.000000	0.003009	0.002508	0.006185	0.008860	0.004681	0.006687	0.004012	0.006352	0.006352	0.008024	0.008860	0.009696	0.012705	0.0
American	0.003343	0.004012	0.005015	0.002842	0.004346	0.004179	0.004012	0.004012	0.004681	0.005015	0.005349	0.006687	0.006687	0.007690	0.007021	0.007355	0.007690	0.0
Delta	0.004681	0.000000	0.000000	0.003176	0.003009	0.003343	0.003678	0.004012	0.004012	0.004346	0.004012	0.004346	0.004012	0.004346	0.005684	0.005684	0.006687	0.0
Frontier	0.011702	0.009696	0.000000	0.003678	0.005015	0.005349	0.007355	0.006018	0.006520	0.007690	0.008024	0.008693	0.009361	0.008693	0.009361	0.009696	0.010030	0.0
Hawaiian	0.000000	0.002340	0.000000	0.000000	0.000000	0.002675	0.002508	0.003678	0.003343	0.002675	0.009027	0.000000	0.000000	0.000000	0.000000	0.004012	0.005015	0.0
JetBlue	0.013708	0.000000	0.000000	0.005349	0.004514	0.003009	0.006185	0.007355	0.006352	0.005349	0.006687	0.008358	0.008358	0.009027	0.012370	0.012872	0.013039	0.0
Southwest	0.000000	0.000000	0.000000	0.002006	0.002340	0.002675	0.002675	0.003176	0.003009	0.003343	0.004012	0.004681	0.005015	0.005349	0.005349	0.006018	0.006687	0.0
Spirit	0.004346	0.025075	0.000000	0.003343	0.004346	0.004681	0.004012	0.004681	0.005015	0.005349	0.006018	0.007021	0.007690	0.008693	0.008024	0.007355	0.009361	0.0
United	0.006352	0.012370	0.000000	0.006018	0.003176	0.004012	0.003678	0.003678	0.005015	0.003845	0.005684	0.005349	0.006352	0.005349	0.005684	0.004346	0.005684	0.0

Airline - Delay Numbers

Airline	Flights	Delays	Delays > 15 mins	Delays > 30 mins	% Delayed Flights	% Delays > 15 mins	% Delays > 30 mins
SOUTHWEST	326028	205677	102730	58451	63%	32%	18%
JETBLUE	41326	20859	13264	9572	50%	32%	23%
FRONTIER	78412	39260	25331	17476	50%	32%	22%
SPIRIT	102558	46589	26199	17307	45%	26%	17%
HAWAIIAN	4004	1753	633	355	44%	16%	9%
AMERICAN	472110	203977	113211	77241	43%	24%	16%
ALLEGIANT	10536	4530	2683	1783	43%	25%	17%
UNITED	231363	92057	44796	30281	40%	19%	13%
DELTA	325740	120978	57268	34946	37%	18%	11%
ALASKA	28598	10530	5255	3328	37%	18%	12%

Day of Week - Kruskal Wallis

```
        Kruskal-Wallis rank sum test

data:  DepDelayMinutes by DayOfWeek
Kruskal-Wallis chi-squared = 458.34, df = 6, p-value < 2.2e-16
```

Comparison of x by group
(Bonferroni)

Col Mean- Row Mean	1	2	3	4	5	6
2	8.802272 0.0000*					
3	-4.818810 0.0000*	-13.29485 0.0000*				
4	-7.383600 0.0000*	-16.01549 0.0000*	-2.365701 0.1890			
5	-10.83614 0.0000*	-19.48971 0.0000*	-5.655128 0.0000*	-3.353712 0.0084*		
6	-5.509680 0.0000*	-13.94645 0.0000*	-0.687698 1.0000	1.652575 1.0000	4.922247 0.0000*	
7	-3.359063 0.0082*	-12.13936 0.0000*	1.582028 1.0000	4.075164 0.0005*	7.508606 0.0000*	2.287369 0.2328

Time of Day - Kruskal Wallis

```
        Kruskal-Wallis rank sum test

data:  DepDelayMinutes by DepTimeBlk
Kruskal-Wallis chi-squared = 12045, df = 18, p-value < 2.2e-16
```