

BUDT 733 – Data Analysis for Decision Makers

Fall 2007 - Sections DC01

Exploring A Trip to Hawaii

What drives flight delays between DC and Honolulu

by

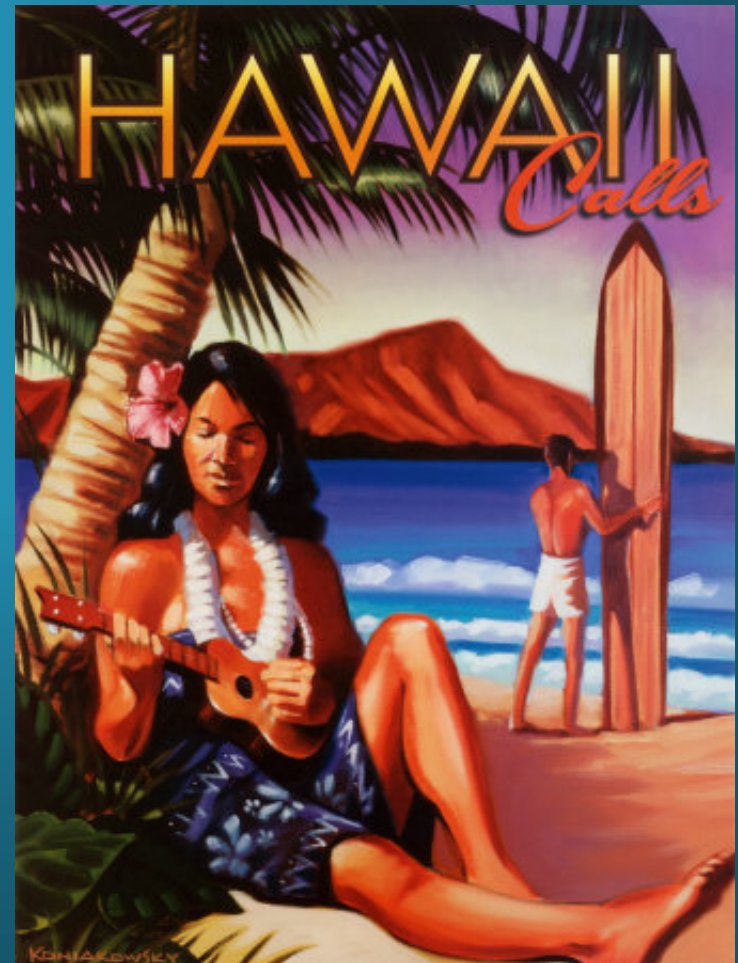
Prashant Bhaip

Andres Garay

Vedat Kaplan

Greg Vinogradovg

Yu-Ling Wang



The Question:

What drives flight delays between DC and Honolulu?

- Explanatory vs. Predictive ?

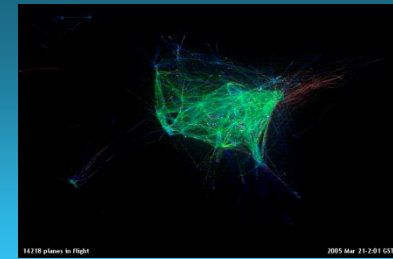


Data

- Expedia to Excel
 - Flights in 2008
- Period: 2004 Q3 – 2007 Q3
- BTS (FAA)
 - 225 MB each Month
 - Total 8.1 GB of Data
- Categorical & numerical






Data Details



DELAY	Weekend	Holiday	OilPrice	MONTH	CARRIER	DAY_OF_VORIGIN	DEST	ORIGIN_T	DEST_VIS	DJ_Close
No	1	0	76.70	9	UA	7 IAD	SFO	77.7	10	13113.38
No	1	0	76.70	9	UA	7 BWI	ORD	75.5	10	13113.38
Yes	1	0	76.70	9	UA	6 IAD	SFO	81.1	9.5	13113.38
No	1	0	76.70	9	UA	6 BWI	ORD	78.3	10	13113.38
No	0	0	76.70	9	UA	5 IAD	SFO	81.1	10	13113.38
No	0	0	76.70	9	UA	5 BWI	ORD	78.8	8.3	13113.38
No	0	0	76.30	9	UA	4 IAD	SFO	80.5	9.7	13363.35
Yes	0	0	76.30	9	UA	4 BWI	ORD	77.7	8.7	13363.35
No	0	0	75.73	9	UA	3 IAD	SFO	75.9	10	13305.47
No	0	0	75.73	9	UA	3 BWI	ORD	74.7	8.8	13305.47
No	0	0	75.05	9	UA	2 IAD	SFO	70.8	9.7	13448.86
No	0	0	75.05	9	UA	2 BWI	ORD	71.4	9.8	13448.86
No	1	0	81.66	9	UA	7 IAD	SFO	61.9	10	13895.63
No	1	0	81.66	9	UA	7 BWI	ORD	63.3	10	13895.63
No	0	1	74.04	9	UA	1 IAD	SFO	75.1	10	13357.74
Yes	0	1	74.04	9	UA	1 BWI	ORD	72	10	13357.74
No	1	0	81.66	9	UA	6 IAD	SFO	65.4	9.9	13895.63
No	1	0	81.66	9	UA	6 BWI	ORD	65.7	10	13895.63
Yes	0	0	81.66	9	UA	5 IAD	SFO	71	10	13895.63
No	0	0	81.66	9	UA	5 BWI	ORD	72.1	10	13895.63
No	0	0	82.88	9	UA	4 IAD	SFO	77.8	10	13912.94

Sources

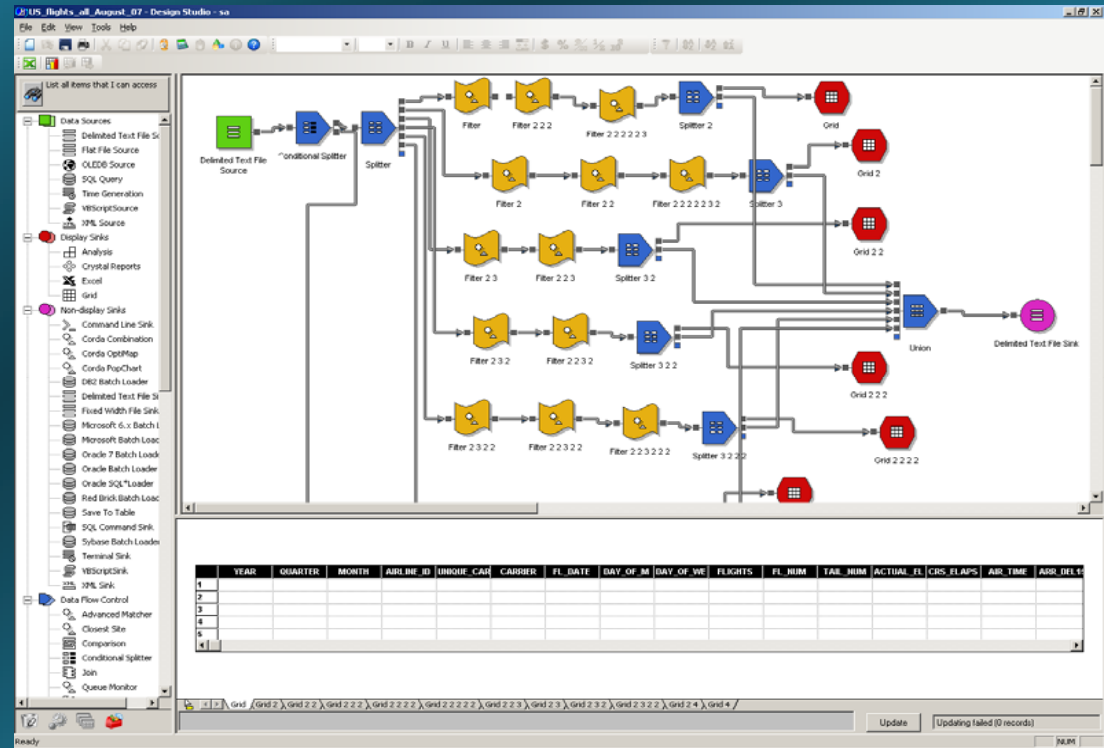
	Departure Time	Origin to Hub	Hub to Destination
 United	9:30 AM	IAD - SFO	SFO - HNL
 Continental	8:00 AM	BWI - IAH	IAH - HNL
 Northwest	8:45 AM	DCA - MSP	MSP - HNL
 Airways	9:30 AM	BWI - ORD	ORD - HNL
 Delta	7:05 AM	BWI - ATL	ATL - HNL
 American Airlines	12:50 AM	DCA - DFW	DFW - HNL

- Expedia
- NOAA (National Oceanic Atmospheric Administration)
- BTS (Bureau of Transportation Statistics)
- Yahoo
- Bloomberg
- Others

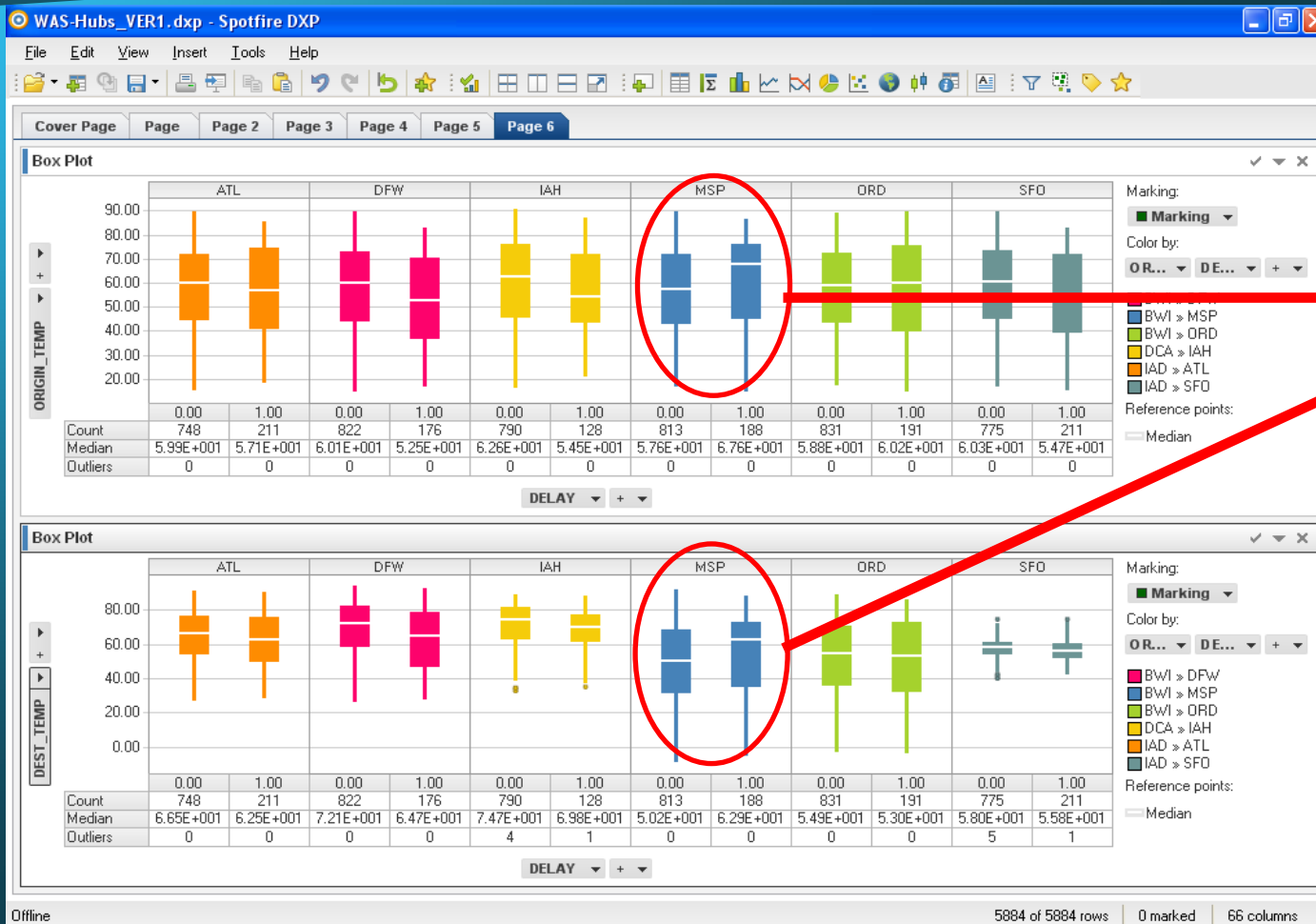


Pre-Processing

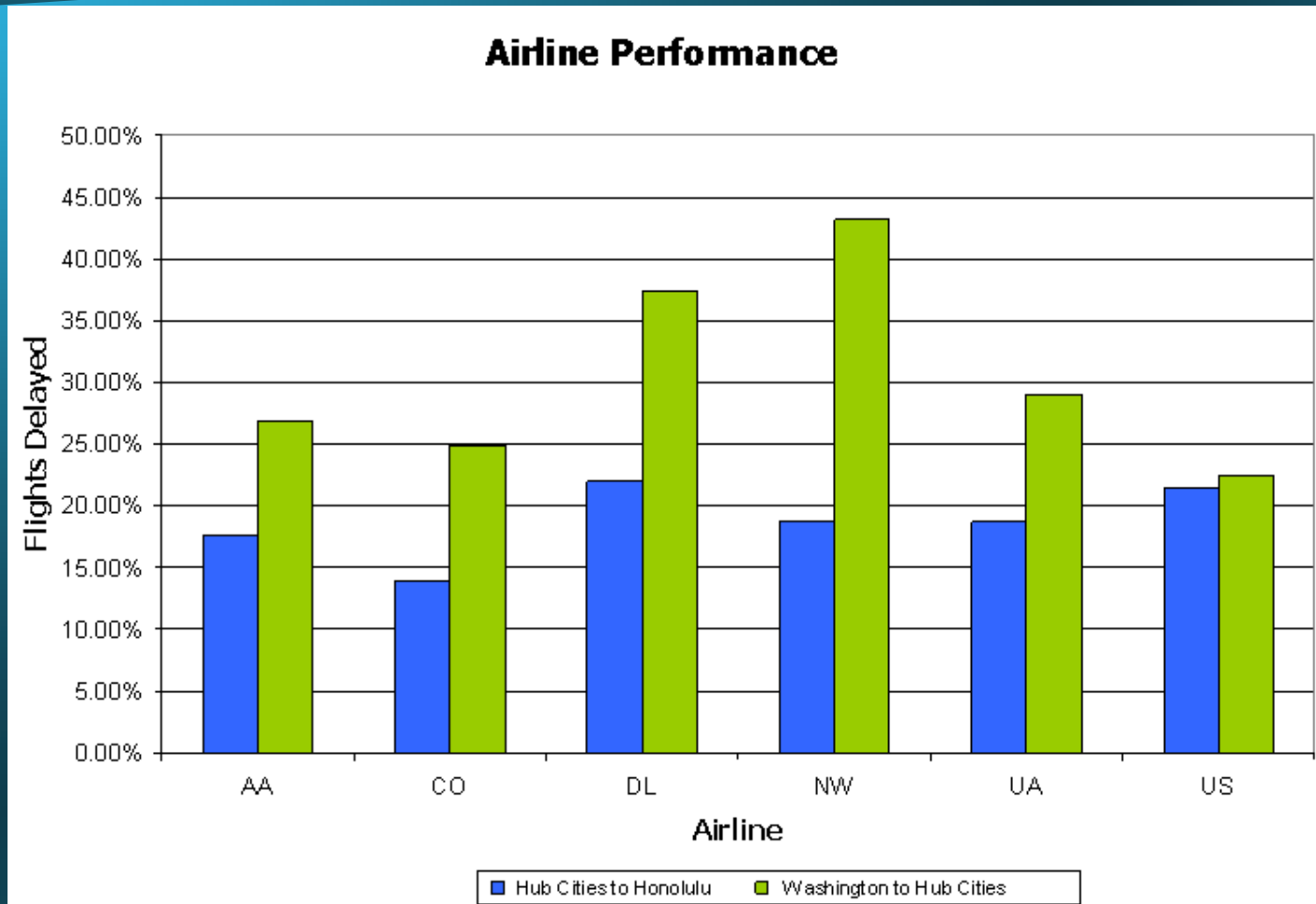
- MM of obs (8.1 GB)
 - Down to 5,000
 - <7MB
- Variables down
 - 100 -> 30 -> 13



Initial Exploration



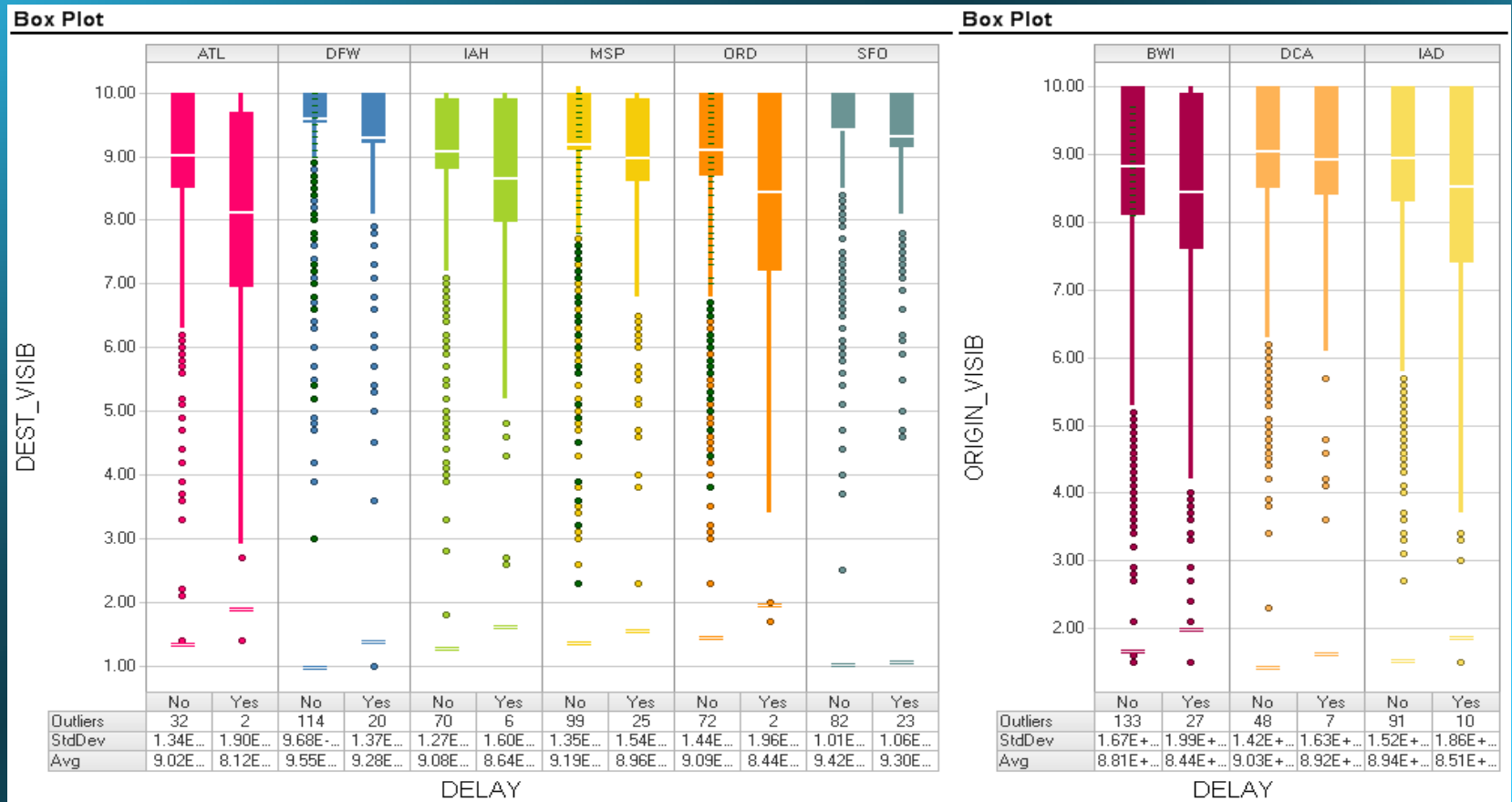
Further Exploration



Impact of Weather

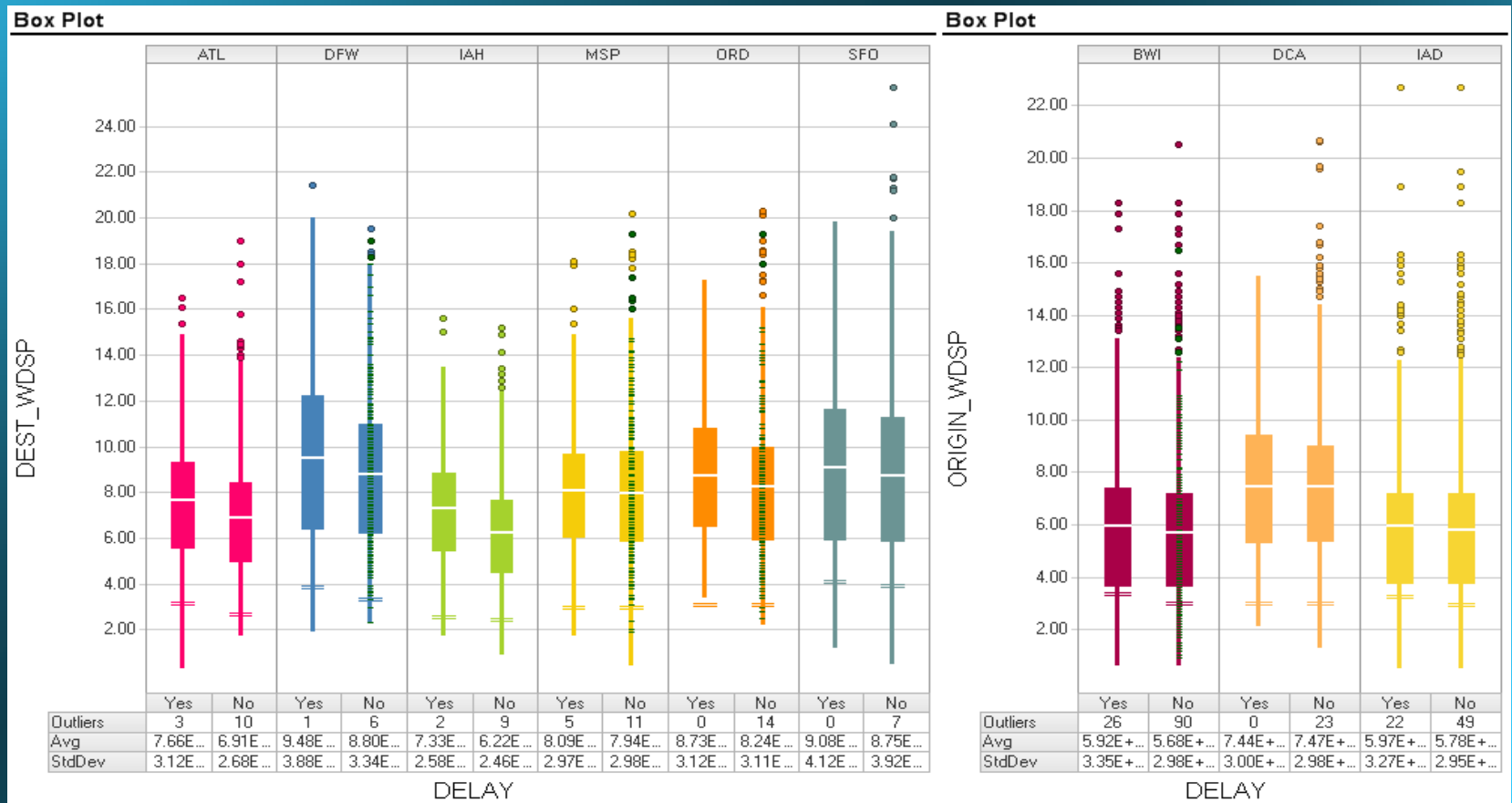
The Visibility Effect

- Adverse visibility conditions cause delays

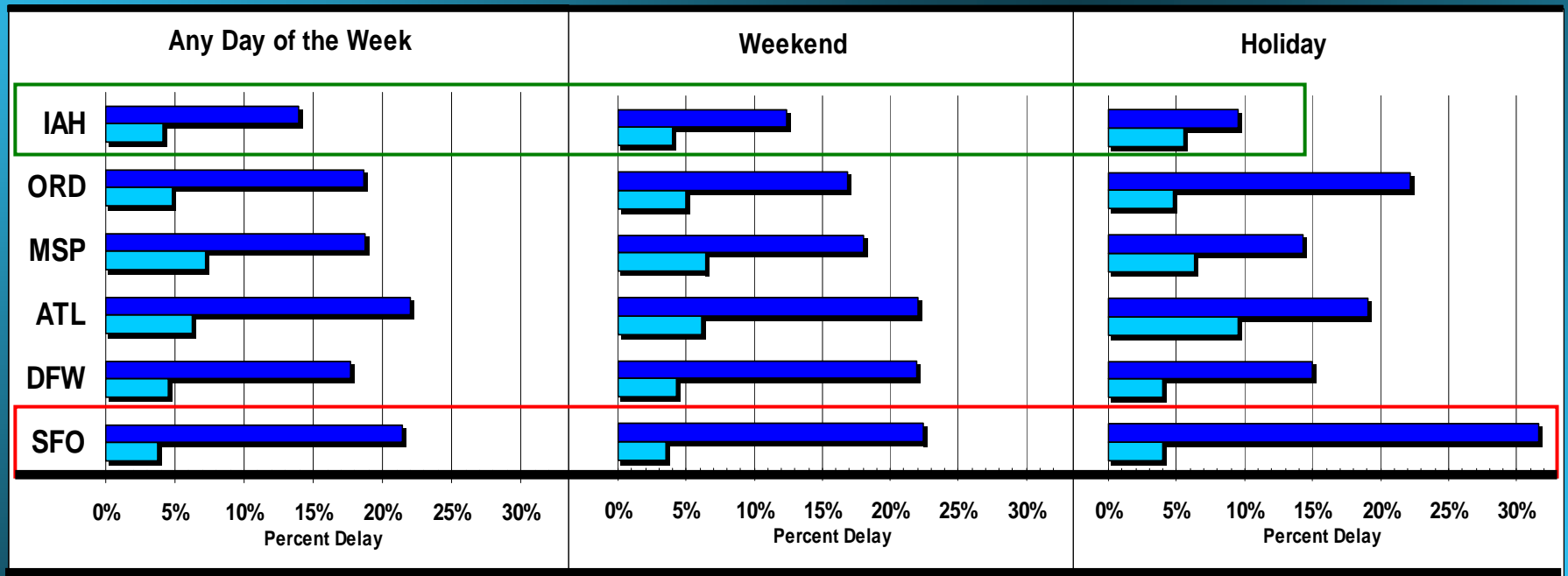


Impact of Weather – The Wind Speed (Gust) Effect

- Wind speed affects takeoff and landing



Impact of the Day of Flight



 DC to Hubs

 Hubs to Honolulu

Does Distance Matter?

- Only the Extreme ones are affected

Washington to Hubs:

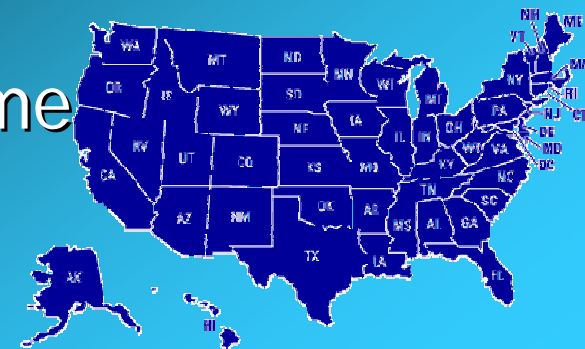
		DEST					
DELAY	Data	ATL	DFW	IAH	MSP	ORD	SFO
No	Count of DELAY	78.00%	82.36%	86.06%	81.22%	81.31%	78.60%
	Average of DISTANCE	533	1217	1208	936	622	2419
Yes	Count of DELAY	22.00%	17.64%	13.94%	18.78%	18.69%	21.40%
	Average of DISTANCE	533	1217	1208	936	622	2419

Hubs to Honolulu:

		ORIGIN					
DELAY	Data	ATL	DFW	IAH	MSP	ORD	SFO
No	Count of DELAY	62.65%	73.15%	75.08%	56.83%	70.99%	77.57%
	Average of DISTANCE	4502	3784	3904	3972	4243	2398
Yes	Count of DELAY	37.35%	26.85%	24.92%	43.17%	29.01%	22.43%
	Average of DISTANCE	4502	3784	3904	3972	4243	2398

Fitting a Model to Understand Delays

- Fitted Classification Tree, Logistic Regression and Discriminant Analysis
 - from DC to the hub cities and then to Honolulu.
- Three techniques yielded similar results
- Partitioned only for pruning the classification tree
- Predictors were dropped gradually
 - p-value
 - overall error of the model
- The R-squared metric showed some improvement over the Naïve rule



Results - Conclusions

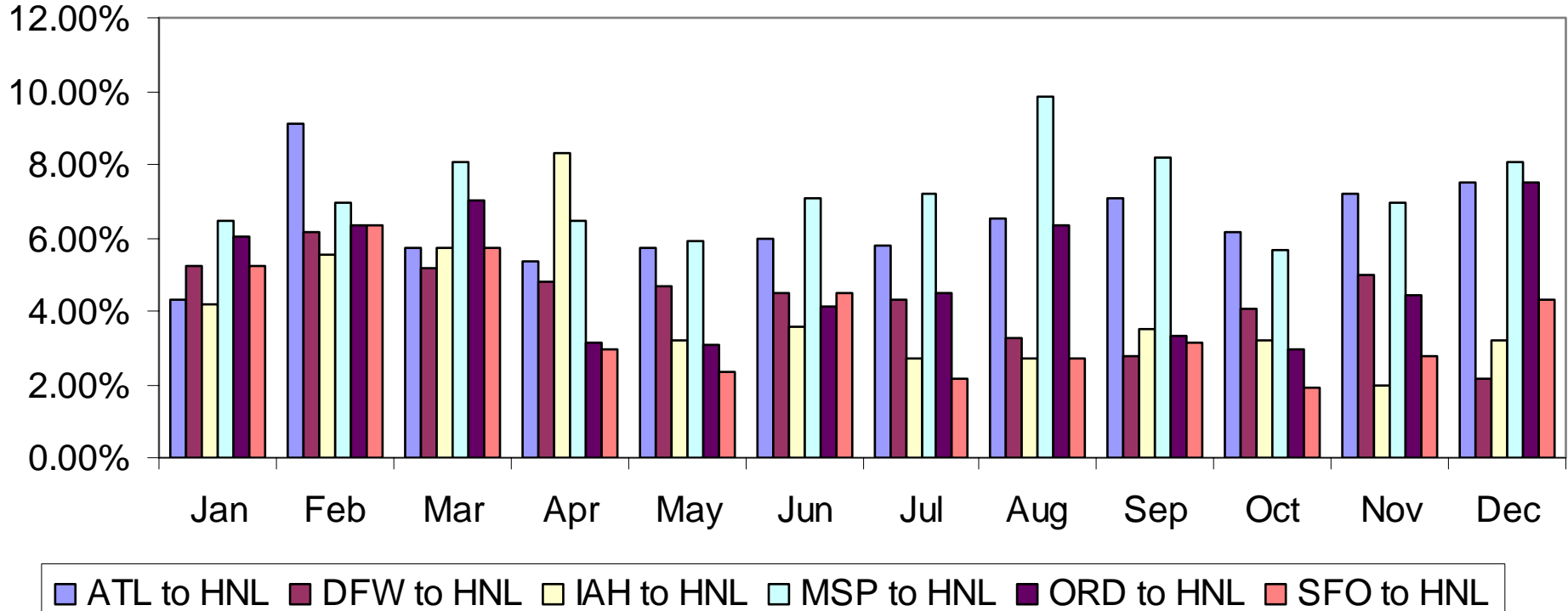
- Weather can cause delay
 - Visibility during landing
 - Wind gusts during both takeoff and landing
- Proxy variables didn't provide good predictors
- Continental is the only airline that showed a significant difference over the others
- Minneapolis handles the winter better than other airports, but cannot handle the additional demand of summer
- Seasonal effects (weekend/holiday) aren't as bad as the media seems to portray

Questions?

Extra Slides

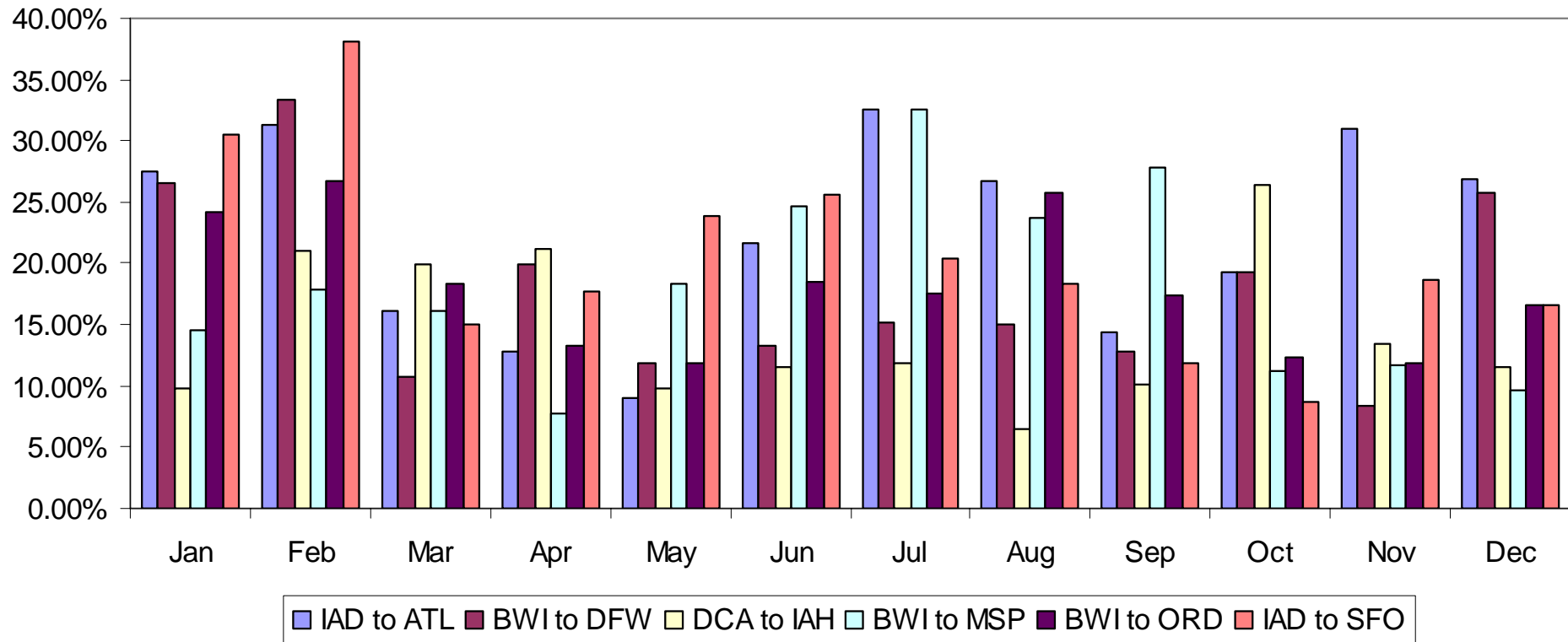
Monthly Delays (Hubs to HNL)

Hubs to HNL - Monthly Delays



Monthly Delays (Was to Hubs)

Was to Hubs - Monthly Delays



Impact of Weather – The Wind Speed Effect

- Wind speed affects takeoff and landing

		ORIGIN			
DELAY	Data	BWI	DCA	IAD	Grand Total
No	Count of DELAY	81.63%	86.06%	78.30%	81.22%
	Average of DEST_WDSP	8.328791565	6.219873418	7.845961917	7.826302574
Yes	Count of DELAY	18.37%	13.94%	21.70%	18.78%
	Average of DEST_WDSP	8.750990991	7.33046875	8.368483412	8.440361991

		DEST						
DELAY	Data	ATL	DFW	IAH	MSP	ORD	SFO	Grand Total
No	Count of DELAY	78.00%	82.36%	86.06%	81.22%	81.31%	78.60%	81.22%
	Average of DEST_WDSP	6.905748663	8.803649635	6.219873418	7.941574416	8.237906137	8.753419357	7.826302574
	StdDev of DEST_WDSP	2.675623832	3.337453837	2.463955106	2.979597037	3.109673201	3.92349124	3.257729443
	Max of DEST_WDSP	19	19.5	15.2	20.2	20.3	25.7	25.7
	Min of DEST_WDSP	1.7	2.3	0.9	0.4	2.2	0.5	0.4
Yes	Count of DELAY	22.00%	17.64%	13.94%	18.78%	18.69%	21.40%	18.78%
	Average of DEST_WDSP	7.658767773	9.478409091	7.33046875	8.089361702	8.731937173	9.078199057	8.440361991
	StdDev of DEST_WDSP	3.122705847	3.878304456	2.582819024	2.969307094	3.116775648	4.123140109	3.456926969
	Max of DEST_WDSP	16.5	21.4	15.6	18.1	17.3	19.8	21.4
	Min of DEST_WDSP	0.3	1.9	1.7	1.7	3.4	1.2	0.3

Logistic Regression

■ Washington to hub airports

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-0.76728678	0.45851296	0.09424377	*
Spring	-0.27423778	0.09222132	0.0029424	0.76015127
Summer	0.38173604	0.11918207	0.0013602	1.46482539
CO	-0.4557862	0.10894412	0.00002868	0.63394934
DISTANCE	0.00012887	0.00005476	0.01860245	1.00012887
ORIGIN_TEMP	-0.0123981	0.00295312	0.00002689	0.98767847
ORIGIN_VISIB	-0.10999046	0.01980485	0.00000003	0.89584267
ORIGIN_WDSP	0.04607463	0.01172593	0.0000852	1.04715252
DEST_VISIB	-0.16257669	0.02505603	0	0.84995091
DEST_WDSP	0.05318914	0.01050471	0.00000041	1.05462909
Dest_Rain	0.34892815	0.08734048	0.00006469	1.41754735
Dest_Thunder	0.3560597	0.11413761	0.00181118	1.42769277
DJ_AdjClose	0.00012257	0.00003183	0.00011775	1.00012255

■ Hub airports to Honolulu

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	99.51849365	14.11064816	0	*
Summer	0.19564669	0.0889582	0.02785587	1.21609712
ATL	0.44447181	0.07911266	0.00000002	1.55966616
MSP	0.76633286	0.08105181	0	2.15186071
ORIGIN_TEMP	0.01606891	0.0057089	0.00488212	1.01619875
ORIGIN_DEWP	-0.01773611	0.00594038	0.00282943	0.98242027
ORIGIN_VISIB	-0.15202869	0.02492423	0	0.85896361
ORIGIN_WDSP	0.02346477	0.00912913	0.01016053	1.0237422
Origin_Rain	0.20867051	0.07800181	0.00746845	1.23203897
Origin_Thunder	0.37560391	0.10014928	0.00017652	1.45587039
DEST_TEMP	-0.03867652	0.01364659	0.00459475	0.96206188
DEST_DEWP	-0.03638766	0.01160606	0.00171719	0.96426642
DEST_SLP	-0.09279958	0.01381092	0	0.91137612

Discriminant Analysis

Variables	Classification Function		
	Delay	No Delay	Difference
Constant	-16676.1523	-16675.832	-0.3203125
Fri_Sun	-10.7436981	-10.799962	0.05626392
Winter	15.32785225	15.27463245	0.0532198
Spring	104.0936813	104.3424377	-0.2487564
Summer	17.5315876	17.21523666	0.31635094
UA	16.2675972	16.31866264	-0.05106544
DL	31.13571739	30.92464066	0.21107673
AA	28.03342819	28.17245865	-0.13903046
CO	-15.2112618	-14.7413464	-0.46991539
DISTANCE	0.00994605	0.00979083	0.00015522
ORIGIN_TEMP	-1.62741184	-1.62127507	-0.00613677
ORIGIN_DEWP	7.64395237	7.64892387	-0.0049715
ORIGIN_SLP	32.22281265	32.222332	0.00048065
ORIGIN_VISIB	-4.29094505	-4.17161465	-0.1193304
ORIGIN_WDSP	28.16441154	28.11614227	0.04826927
Origin_Fog	-3.49710989	-3.61216617	0.11505628
Origin_Rain	51.66131973	51.71683502	-0.05551529
Origin_Snow	107.2143326	106.9103241	0.30400848
Origin_Thunder	11.14117813	11.07869434	0.06248379
Origin_Tornado	-66.7338486	-68.5454712	1.81162262
DEST_TEMP	1.57903779	1.58504903	-0.00601124
DEST_DEWP	-3.18011236	-3.1888907	0.00877834
DEST_SLP	0.01533218	0.01517634	0.00015584
DEST_VISIB	13.62252235	13.79381847	-0.17129612
DEST_WDSP	0.7925173	0.73128134	0.06123596
Dest_Fog	0.2684176	0.19592339	0.07249421
Dest_Rain	11.01359844	10.70842648	0.30517196
Dest_Snow	75.68932343	75.75340271	-0.06407928
Dest_Hail	-29.5155144	-30.0580025	0.5424881
Dest_Thunder	-3.43515801	-3.79444313	0.35928512
DJ_AdjClose	0.00243849	0.00230125	0.00013724

Variables	Classification Function		
	1	0	Difference
Constant	-1681.75708	-1687.8158	6.05871582
Fri-Sun	6.65679216	6.67501497	-0.01822281
Weekend	-1.88453269	-1.80504501	-0.07948768
Holiday	7.42878485	7.33501244	0.09377241
Winter	94.89620209	94.92543793	-0.02923584
Spring	67.32580566	67.34754181	-0.02173615
Summer	-16.3136196	-16.5240231	0.21040345
UA	17.21924591	17.98859024	-0.76934433
DL	6.80307913	7.21482086	-0.41174173
AA	12.43914986	13.21653271	-0.77738285
CO	16.18800545	17.13852501	-0.95051956
DISTANCE	0.01421429	0.01409223	0.00012206
ORIGIN_TEMP	-0.17978559	-0.18232796	0.00254237
ORIGIN_VISIB	8.62186432	8.72542095	-0.10355663
ORIGIN_WDSP	0.87387049	0.8526516	0.02121889
Origin_Fog	13.8432188	13.59267902	0.25053978
Origin_Rain	10.81028461	10.68365479	0.12662982
Origin_Snow	15.27459908	15.09401608	0.180583
Origin_Hail	-23.5604248	-23.3107739	-0.24965095
Origin_Thunder	-9.97609425	-10.3168039	0.34070968
DEST_TEMP	17.85284615	17.92337418	-0.07052803
DEST_VISIB	176.6824188	176.8218994	-0.13948059
DEST_WDSP	-5.29001379	-5.27540874	-0.01460505
Dest_Fog	138.6786041	138.4454041	0.23320008
Dest_Rain	36.43346024	36.54920959	-0.11574935
Dest_Thunder	102.8440552	102.8561783	-0.0121231
Dest_Tornado	-48.5666733	-48.637085	0.07041168
DJ_AdjClose	0.00722685	0.00709572	0.00013113

Classification Tree

