A2. Exploratory Data Analysis (Airline Data Set)

Greg Ceccarelli | February 4th, 2016

W209 - 4

Hypothesis 1: The larger the airline (as measured by aggregate flight count per day/week/month) the worse (higher) the average elapsed flight minutes per mile flown in the September 2001.

Beginning: View 1

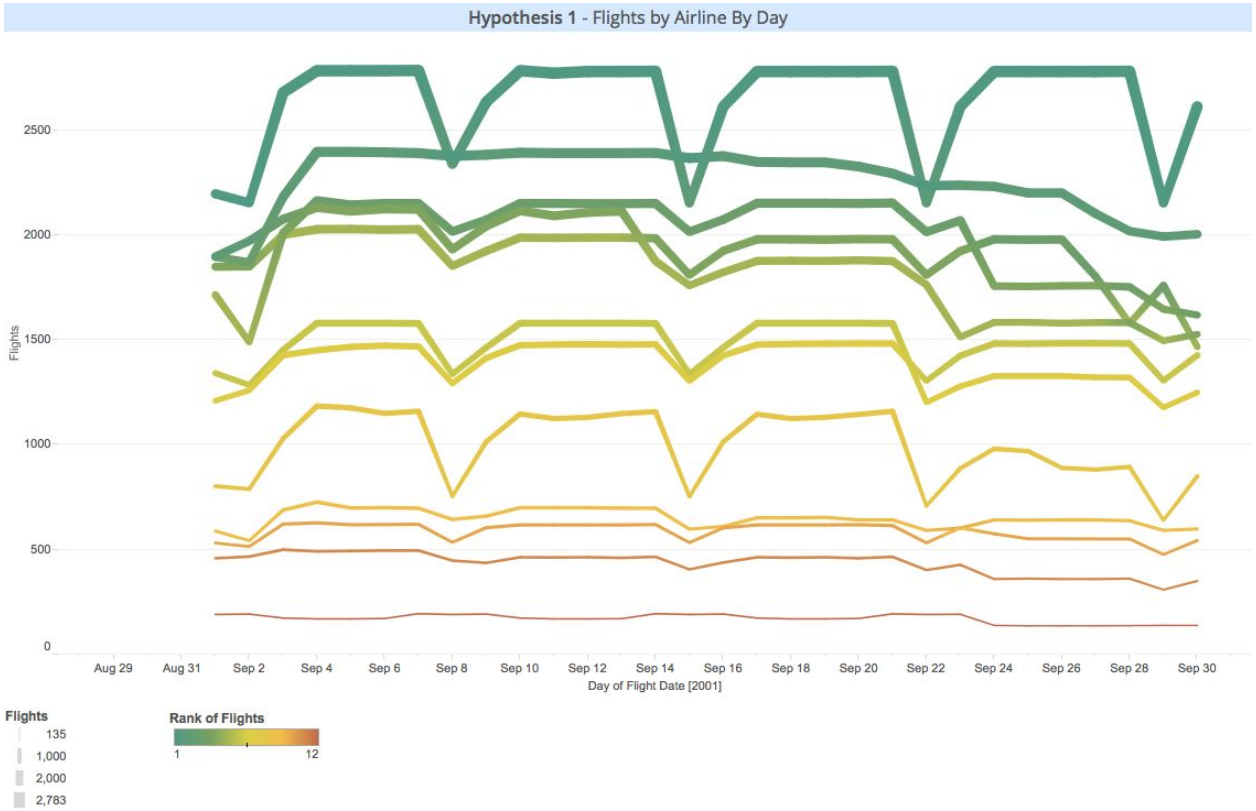
Hypothesis 1 - Flight Volume & Airline Rank

Carrier Name	Airline Volume Rank	Flights
Southwest Airlines Co.	1	79,079
Delta Airlines	2	67,821
United Airlines	3	59,590
US Airways	4	57,377
American Airlines	5	55,846
Northwest Airlines	6	43,275
Envoy Air Inc.	7	42,811
Continental	8	29,893
T'Way Air Co., Ltd.	9	19,427
Amapola Flyg AB	10	17,495
Alaska Airlines	11	13,002
9 Air Co Ltd	12	5,082

What's informative about this view: This table shows the total count of scheduled flights in the month of September 2001. This table is very helpful to understand baseline airline size as measured by their total scheduled flight count in the period. I've decoded the carrier codes into the actual airline names during the period and ranked the volume.

What could be improved about this view: At current this data is very aggregated and gives no context with regard to 1) flights that actually occurred during the period or 2) their performance. Additionally, a way to view this volume over time would be useful

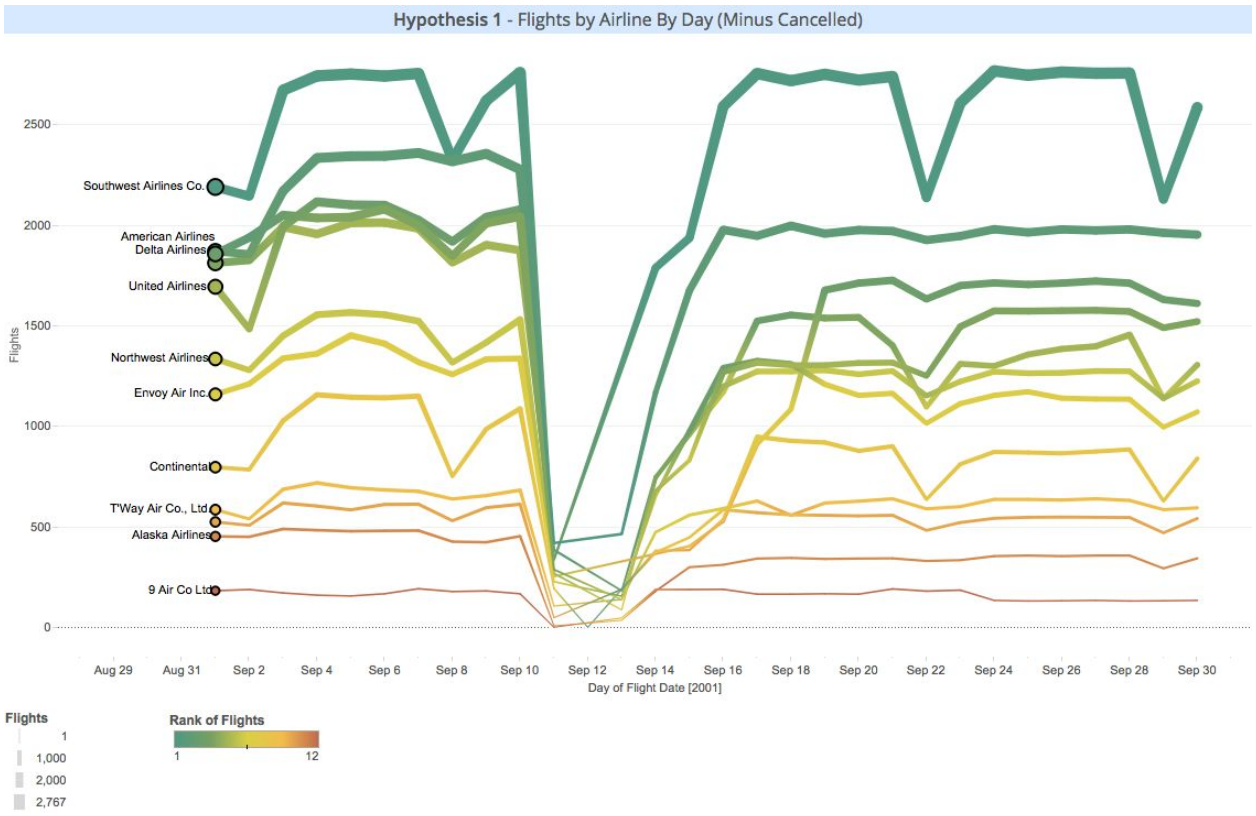
View 2



What's informative about this view: This view provides more detail into the flight operations of each airline and their rank over time. It gives you a better visual sense of scale of each airline's operation in comparison with each other. Additionally, viewing the data over time enables one to understand pertinent details that separate the carriers from one another. For example, Southwest's volume drops off over the weekend (catering mostly to consumers) while Delta has much stabler volume (picking up after the first week of the month) due to their target market (business customers).

What could be improved about this view: We still don't have an understand of elapsed flight time performance and there seems to be something fishy about September 11 - 18th flight volume. I would have expected there to be a significant dropoff here given the contextual knowledge of the events surrounding that time period.

View 3



What's informative about this view: Wow look at that precipitous drop off of flights! Given the volume drops, it's very interesting to review the latency with which each airline is able to get back up and running to their pre 9/11 operational business as usual.

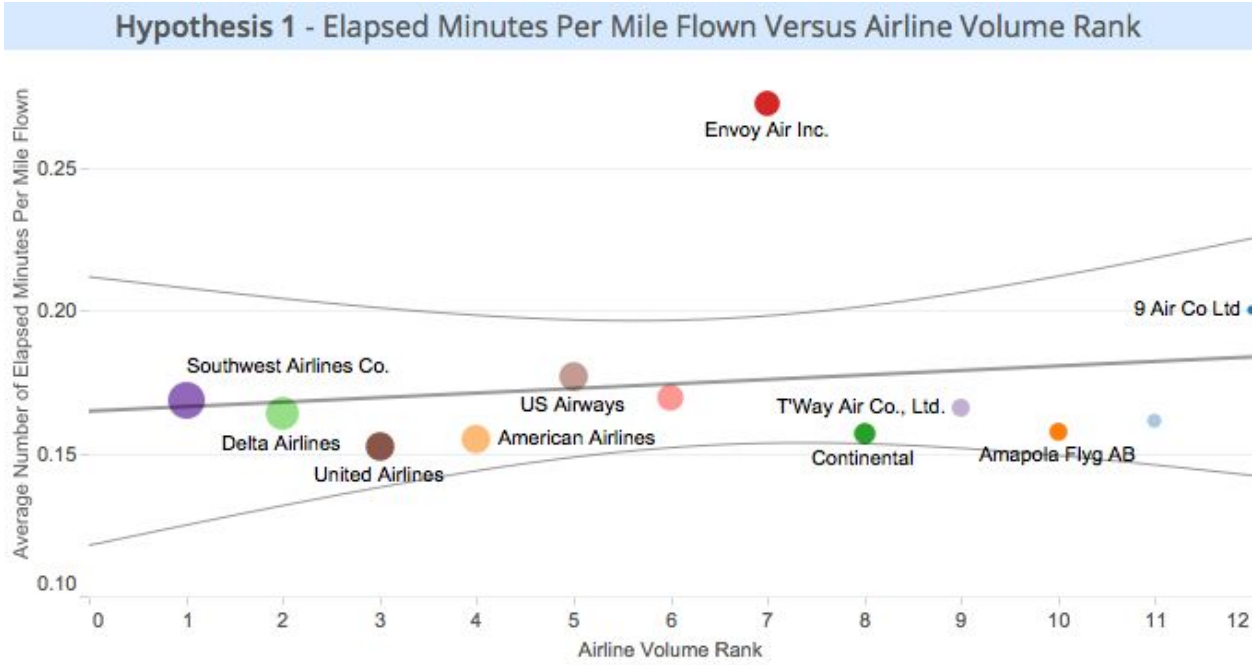
What could be improved about this view: Again, we still need to weave in average flight performance somehow to show whether or not we can confirm the hypothesis.

View 4

What's informative about this view: This view shows and encodes a lot of information. First it shows, on the Y Axis, Average Number of Elapsed Minutes per Mile Flown. We see that across the board, the average number of minutes per mile ~.17. Additionally, volume rank (by Day) is still encoded with color. We see Southwest (highest aggregate flight volume) is ranked number one while 9 Air Co is ranked number 12 in terms of volume. What's interesting here, is that some of the large other carriers including United and American Airlines have the lowest average number of elapsed minute per mile flown.

What could be improved about this view: This view does contain a lot of information and could possibly be simplified. Perhaps by changing the way the volume or rank is encoded could improve interpretability.

Final: View 5



What's informative about this view: Finally a simplification of the data that enables one to quickly grok the required information. This view shows quite clearly how airlines rank in terms of volume versus how efficient they are operating in the air.

What could be improved about this view: I think this final simplification provides the requisite information in order to confirm/deny my original hypothesis. It quite clearly shows a (slightly) positive correlation between airline rank and Minute Per Mile which disproves my original hypothesis

----

Conclusion These data do not confirm my original hypothesis that “The larger the airline (as measured by aggregate flight count per day/week/month) the worse (higher) the average elapsed flight minutes per mile flown in the September 2001”. Contrary to my original belief, the data would actually suggest that the larger the airlines’ operation, the more efficient they are (in aggregate). The data clearly show that some very small airlines are incredibly inefficient in the air which may be a by product of their fleet (size of planes / engine speed, etc) or other variables such as being subject to in air holds, etc.

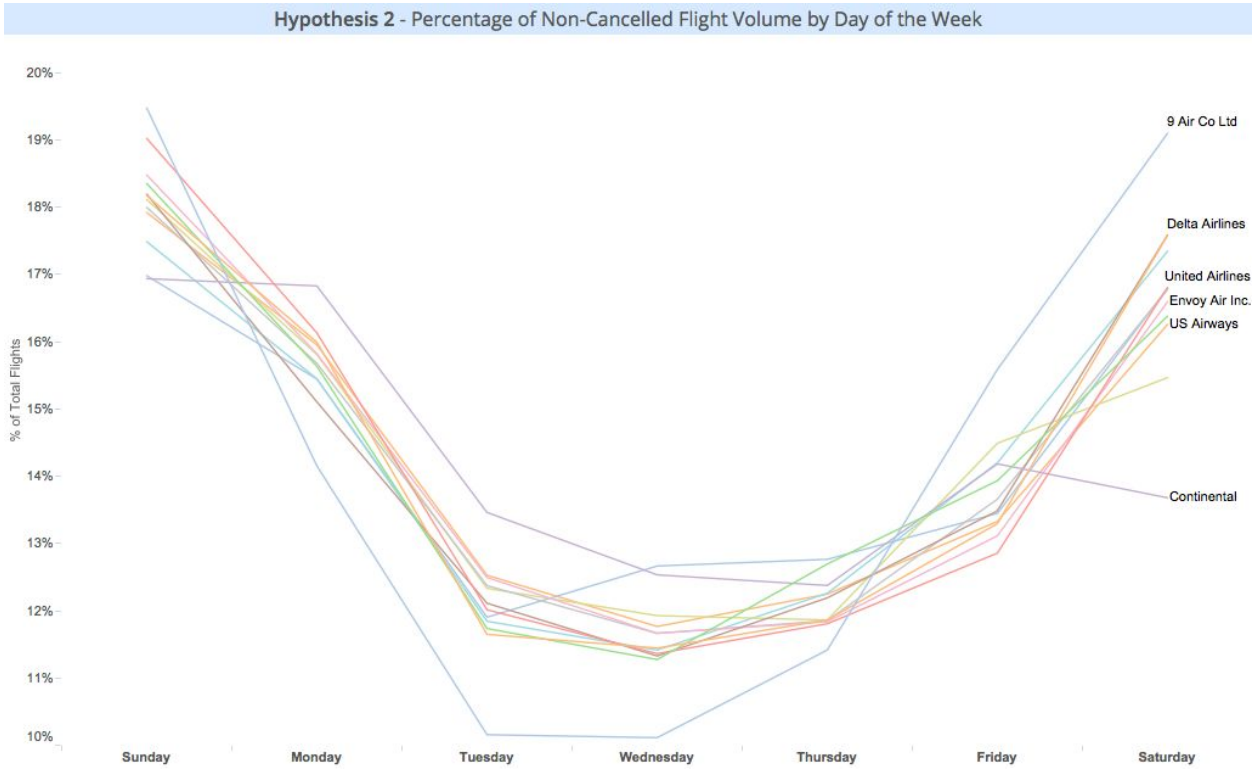
Hypothesis 2: There are noticeable differences in operational scheduling by airline. Said differently, different airlines have different flight volumes by day of the week to cater to different audiences.

Beginning View

Hypothesis 2 - Percentage of Non-Cancelled Flight Volume by Day of the Week

Carrier Name	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
9 Air Co Ltd	19%	14%	10%	10%	11%	16%	19%
Alaska Airlines	18%	16%	12%	11%	12%	13%	18%
Amapola Flyg AB	18%	16%	12%	11%	13%	14%	16%
American Airlines	19%	16%	12%	11%	12%	13%	17%
Continental	17%	17%	13%	13%	12%	14%	14%
Delta Airlines	18%	15%	12%	11%	12%	13%	18%
Envoy Air Inc.	18%	16%	12%	12%	12%	13%	17%
Northwest Airlines	18%	16%	12%	12%	12%	14%	17%
Southwest Airlines Co.	18%	16%	12%	12%	12%	14%	15%
T'Way Air Co., Ltd.	17%	15%	12%	11%	12%	14%	17%
United Airlines	17%	15%	12%	13%	13%	13%	17%
US Airways	18%	16%	13%	12%	12%	13%	16%

Final View



Hypothesis 3: Airline volume after 9/11 didn't pick up to operational norms for at least a week after the event

Beginning View

Hypothesis 3 - Day over Day percentage difference in flight volume post 9/11

	2001									
	Q3									
	September									
Carrier Name	10	11	12	13	14	15	16	17	18	19
9 Air Co Ltd	168	1		48	189	189	190	166	166	168
Alaska Airlines	454	8		36	181	300	312	343	346	341
Amapola Flyg AB	613	49		190	373	448	586	571	560	558
American Airlines	2,043	288		139	679	831	1,270	1,316	1,304	1,304
Continental	1,088	194	1		382	385	538	948	928	920
Delta Airlines	2,277	420		465	1,168	1,675	1,977	1,947	1,997	1,958
Envoy Air Inc.	1,337	228		156	655	979	1,291	1,326	1,310	1,209
Northwest Airlines	1,531	269		88	742	970	1,199	1,273	1,273	1,279
Southwest Airlines Co.	2,759	337			1,788	1,936	2,591	2,754	2,718	2,750
T'Way Air Co., Ltd.	683	107		140	474	559	592	629	558	619
United Airlines	2,077	254			366	405	527	911	1,085	1,677
US Airways	1,875	386		183	748	957	1,170	1,524	1,554	1,538

Final View