7-11-2013

The Effect of Airport Delays on the Evolution of the U.S. Air Travel Network

Eric Cox

Eric Cox
Candidate

Geography
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Constantine Hadjilambrinos , Chairperson

Karl Benedict

Scott Freundschuh

THE EFFECT OF AIRPORT DELAYS ON THE EVOLUTION OF THE U.S. AIR TRAVEL NETWORK

by

ERIC COX

Bachelor of Science, Computational Mathematical Sciences
Arizona State University
2007

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science in Geography

The University of New Mexico
Albuquerque, New Mexico

May 2013

Acknowledgements

I would like to thank all of my committee members for their frequent help and occasional understanding during this process. Karl, your feedback was invaluable. Maria Lane deserves a lot of credit for convincing me of the wisdom of writing a thesis in the first place and for getting me off to a very good start. I would like to thank my wife her forceful encouragement to finish. And most of all I need to thank, Mike Abernathy my employer and friend who has generously sponsored my studies and thereby enabled me to see them through.

The Effect of Airport Delays on the Evolution of the U.S. Air Travel Network

by

Eric Cox

B.S., Computational Mathematical Sciences, Arizona State University, 2007
M.S., Geography, University of New Mexico, 2013

ABSTRACT

An investigation is made into the question of how U.S. airlines respond to airport-based delays at domestic airports using data from the FAA's On-Time Performance database and aircraft inventories for major U.S. Airlines. Three delay mitigation techniques are studied: increasing aircraft size, rerouting transit passengers, and decreasing schedule peaking. Regression analysis is used to determine where significant relationships exist between study variables and the overall level of flight delay for all airlines at each airport they serve. T-Tests indicate schedule peaking is more likely to be increased at airports with higher levels of delay, but that no specific airline undertakes this strategically, and that airlines are not more likely to make changes at airports where they are more dominant. However no airlines were found to make any changes at airports where there are no competing airlines.

Table of Contents

List of Figures

List of Tables

Introduction

Congestion-based delays are a major problem at U.S. airports, and one that is likely to get worse. Every delayed flight costs its carrier significantly (Forbes, 2008; Britto, et al., 2012; Ferguson, et al., 2012) as well as inconveniencing, or even stranding its passengers. Furthermore, each delayed flight can impose delays on other flights at an airport leading to a cascade of delays that can ruin hundreds if not thousands of travelers' plans. Several techniques to alleviate congestion exist, but in many cases the only practical option is changing routes and schedules for the airlines that use the airport.

This study examines the extensive archives of data maintained on the subject as time series and across space. Doing so should provide a much better understanding of how airlines have reacted to changes in congestion delays at airports in the U.S. in the past and may point toward better ways to manage the issue in the future.

Background

There are a number of important processes affecting the evolution of the airline industry, the most important of which is growth. The air transportation industry has been growing more rapidly than most other transportation sectors of the economy essentially since its inception (Wilken, et al., 2011). Economic downturns and disasters such as the 9/11 terrorist attacks or the recent recesssion can slow or temporarily reverse this trend (Johnson, 2006), but it remains strong over larger timescales (Swan, 2002). There is some speculation regarding how long such a trend can possibly continue given environmental and technological considerations, but there is generally little reason to imagine that demand will fall below current levels in the next several decades and most experts anticipate strong growth (Pai, 2011; Evans and Schäfer, 2011).

While aircraft can be constructed, replaced or decommissioned in order to meet changes in demand, the airline industry relies heavily on ground-based infrastructure that is much more difficult to scale-up. Commercial flights require terminals, runways, and air traffic control at both the origin and the destination. Runways in particular can be difficult to add because they require a great deal of real estate. For instance, Los Angeles International Airport's *smallest* runway has 1.3 million square feet of tarmac, and this does not include the large buffer zones required on all sides (FAA, 2012). In areas with high population density this quantity of space is not only extremely costly to acquire, but more often than not, is simply not available (Debbage, 2002; Zografos, 2008; Flores-Fillol, 2010). However, highly populated areas are precisely the areas with the greatest demand for air travel. Even if the necessary land is available, these expansions tend to be blocked politically due to their unpopularity with local residents opposed to the increased air and noise pollution they cause and the possible resulting drop in home values (Debbage, 2002; McMillen, 2004). Each runway can only accommodate a limited number of aircraft arrivals or departures per hour, depending on the runway and the mix of aircraft being used. Generally a time buffer of 4-6 minutes is required between flights as a safety precaution (Swan, 2002).

In many European countries these considerations have led national governments and airport managers to strictly limit the number of flights that can be scheduled at most airports and the times for which they can be scheduled (Santos, 2010). In the U.S. the FAA does not favor these types of restrictions and has only allowed them at Chicago O'Hare, La Guardia, JFK (New York), and Washington National. It has imposed and repealed the restrictions a few times and currently none are in place at O'Hare although it is consistently the most congested airport in the country. Several reasons exist to avoid 'slot-restrictions.' They limit airline's flexibility to provide passengers with the most desirable schedule. They require a large bureaucracy to implement them. The costs and benefits of the system are difficult to distribute fairly among stakeholders. The issue of how to distribute slots is also a contentious one (Debbage, 2002). Airlines that gain control of slots may also be tempted to hoard them to prevent competition from other airlines, but there is little evidence of this in practice. (Debbage, 2002).

Another important process in the recent history of the U.S. Airline industry is deregulation. During the late 1970's and early 1980's the FAA deregulated the airline industry thus allowing airlines to schedule or reschedule flights freely, to choose which destinations they would serve and to allow more airlines to enter the market than previously could. This has resulted in lower fares for passengers. Since airlines are now allowed to compete openly across the entire domestic market many of them have also struggled to remain profitable. After deregulation, most existing airlines began to adopt more centralized route networks based on a hub-and-spoke pattern where most flights connect through a limited number of hub airports. This means that there are fewer direct flights, but it also allows the airline to service more destinations with the same number of flights, and to provide more frequent service at most destinations. As a consequence, hub airports become much more congested as they must service a growing number of connecting passengers in addition to travelers who are actually embarking or disembarking there. The problem is further exacerbated by the practice of 'schedule peaking'. In order to make connections easier for passengers at hubs, airlines tend to schedule their arrivals and departures more closely together than otherwise necessary (Daniel and Harback, 2009). This results in large numbers of flights arriving in quick succession followed by a lull followed by large numbers of flights departing in quick succession resulting in much more congestion than if the arrivals and departures were more evenly spaced.

Industry Conventions

The FAA requires commercial airlines to keep extensive records regarding all of their aircraft, expenditures, flights scheduled, etc. Each aircraft (even private, non-commercial aircraft) in the U.S. is assigned an N-Number or Tail Number that uniquely identifies it, similar to an automobile license plate number. It is composed of the letter N, 1-3 numerical digits (not starting with zero) and 0-2 alphabetic letters.

Airports are frequently referred to by their IATA code, which is a three-letter abbreviation assigned to each airport. For airports included in the dataset, these codes, the names of their respective airports and their locations are listed in Appendix D. It is also common in the airline industry to refer to airlines by a two-letter (one number may be used) codes assigned to them by the IATA. For airlines included in the dataset these abbreviations are listed in Appendix E. Both codes are frequently seen on boarding passes, departure/arrival boards and in airline-related research. These codes and the names of the respective airlines for airlines included in the dataset are listed in Appendix E. Both types of codes will be used throughout the text.

Literature Review

Air travel is an integral part of the modern economy and touches all of our lives in many ways, but this convenience depends on the smooth operation of a complex network of airline connections and interactions that constantly evolve in subtle ways. Increasingly this network is affected by the creation and propagation of delays to scheduled flights, many of them caused by the ever-increasing demand on limited runway and air-traffic control resources (Pai, 2010). Much work has been on done on the development of this network, especially in the wake of recent deregulation. A significant amount of literature can be found examining the causes and propagation of delays along with a large number of studies considering alternatives for alleviating or at least slowing the growth of delays. However, very little is known about how the existence and worsening of congestion-based delays may affect the evolution of the nationwide air-travel network and airline schedules.

Global Airline Network Structure and Development

The air-travel network, like other travel networks, is dependent on the existence of a broad array of infrastructure. However, unlike other travel networks, it constantly evolves independently of that infrastructure as airlines modify schedules and routing. After airline deregulation in the late 1970's airlines have been much freer to adopt any type of network structure that they choose. In many cases this led to the adoption of hub-and-spoke networks where traffic from a single airline is concentrated at one or very few airports, where they become the dominant carrier (Reynolds-Feighan, 2001). Conversely it has also lead to the development of so-called 'low-cost carriers', such as Southwest airlines, which tend to operate much more dispersed, fully-connected networks and serve secondary airports (Francis, et al. 2006). A number of techniques have been used to study and quantify these changes. While traditional thinking has favored the idea that post-deregulation networks would become much more concentrated at major hubs, evidence is mixed (Reynolds-Feighan, 2001; Derudder and Witlox, 2009).

The impact of these types of changes has been the focus of many studies in recent years. Cities which become hubs offer considerable advantages to citizens and businesses located nearby. This includes lower fares, shorter travel times more frequent flights and more direct flights all of which increase the accessibility of the city (Grubesic and Zook, 2007). However, increased dominance of a single airline at an airport is shown to increase fares (unless that airline is Southwest) (Van Dender, 2007). Accessibility is generally higher in large urban centers, especially those with more than one major airport, but varies considerably across the United States (Matizsiw and Grubesic, 2010). Residents of more remote areas tend to make greater use of airports and feel more positively about the impact that air travel has on their travel accessibility, despite the fact that they receive less absolute benefit from it (Halpern and Brathen, 2011). Although some governmental programs exist to foster greater accessibility to air travel for remote areas of the U.S., the results tend to be somewhat ineffective (Matisziw and Grubesic 2011).

A great deal of work has been done to examine how network changes have impacted the concentration of traffic (and therefore the creation and intensification of hubs) since deregulation. The overall trend in the U.S. has been for the largest markets to gain traffic while smaller markets lose traffic (Bhadra and Kee, 2008). Several different indices have been used to quantify traffic and airline concentration. Generally GINI or

Herfindahl indices have been preferred for their responsiveness to differing strategies. GINI, Herfindahl, and Theil indices were used to determine that major carriers did adopt more concentrated networks after deregulation and that many low-cost carriers operate less concentrated networks (Reynolds-Feighan, 2001). A variety of indices show that concentration patterns in Europe are found to be somewhat greater, especially among national flagship carriers which operate a large volume of intercontinental flight from national capitals and vary substantially from country to country (Derudder and Witlox 2009; Huber 2009). Herfindahl-Hirschmann indices were used to examine delays at market-concentrated airports (where one airline has a large share of traffic) (Diana, 2009). The Herfindahl-Hirschmann index was found to agree very well with a survey of industry experts regarding which airports are major hubs (Tiago et al., 2010) and is probably the most used. Despite relatively stable demand and network structure, the Nyusten-Dacey method has been used to show that the global hierarchy of airports is extremely unstable, although it is unclear how this reflects the realities of travel and whether it reflects similar shifts in flight concentration (Grubesic, et. al. 2009).

Several studies have attempted to determine what leads an airport to become an airline hub, or a dominant airport. Nash-Equilibria were used to examine the advantages of different hub locations in the South-Atlantic market, noting that while a central location is advantageous, a fairly large destination market is also required to make a hub practical (Martín and Román, 2003). It was also recommended that non-hub airports offer more flights to hub airports in order to become more competitive (Martín and Román, 2003). An explanatory model of air traffic for U.S. airports indicated that the most important factors are local population and the distance to the nearest airport, but that percapita income, tourism, and technical/management employment also have an important role to play (Liu et. al. 2006). An examination of the competition for international transfer passengers revealed that several European airports are well situated geographically to capture transfers due to their ability to serve as a connection between Europe, the U.S. and Asia. However Atlanta had better performance due to the number of connections that it offers (Redondo et. al. 2011).

Finally, a number of papers have sought to compare hub-and-spoke with fully-connected network models. The emergence of low-cost carriers (usually operating fully-connected networks) has sparked controversy. Despite evidence of unfair competition (Dobruszkes, 2006) low cost carriers have provided benefits to passengers such as flying routes traditional airlines no longer fly, providing more seats, and lower ticket prices (Francis et al., 2006). Low cost airlines appear to emerge consistently within a few years of deregulation and become so successful that competition becomes fierce and many airlines fail (Francis et al., 2006). This pattern first appeared in the U.S. and Europe, but may play out in Asia and other developing markets which are still highly regulated (Graham et al. 2006). An examination of European low-cost carriers showed that they are responsible for roughly 50% of the growth in available seats in European markets (Dobruszkes, 2006).

Economic Geography of Airport Capacity

In recent years an ever-increasing demand for air travel has led airlines to demand more arrivals at airports than can easily be accommodated either on the runways or by air-traffic control capabilities (Evans and Schäfer, 2011). In response many airports have—or are planning to—increase these capacities, often at great cost. Arguably there are a number of real benefits to greater capacity and more air traffic, but there are also a number of important costs and drawbacks for the host city to consider as well. These are detailed below. In many cases barriers exist that make these expansions impractical where they are most desired. However, airports that do not expand or offer the level of service or airfare that their competitors do face a real danger of losing passengers (Van Dender, 2007).

There are a number of studies which indicate the important economic benefits of being a city with a high level of air-traffic. There is enough evidence to indicate that air traffic is a causative factor for economic growth—especially employment in the technical and management fields (Button and Lall, 1999; Debbage and Delk, 2001; Brueckner, 2003) as well as on overall employment level and population growth (Green, 2007). Several studies also indicate that having more traffic often means lower fares (Bhadra and Kee, 2008; Grubesic and Zook, 2007). Due to these advantages many local governments have invested in airport infrastructure in order to produce economic growth. It appears that in so doing some airports are able to purchase more air-traffic at about $266 per departure but it is unclear whether this leads to economic growth (Nunn, 2005). The effects of these investments in expansion are often uncertain. Where there is a single dominant airline, capacity expansion may constitute a windfall for the airline without producing lower fares, but at more competitive airports it may increase competition and decrease fares (Fageda and Fernández-Villadangos, 2009).

On the other hand, air traffic also presents a number of drawbacks, such as air and noise pollution. Air traffic noise has a significant depressive effect on real-estate value for areas near the airport—as much as 9.2% in the 57 square miles surrounding O’Hare

(McMillen, 2004). Also, Chicago would likely see less air pollution if it did not complete its planned expansion of the airport (Evans-Schäfer, 2011). Similarly, airport capacity expansions are beginning to meet stiff environmental opposition in Europe, which will likely lead to demand for runway slots in major cities to outstrip supply (Graham and Guyer, 1999).

Airports are facing increasing pressure to improve facilities; especially as internet-based airfare shopping makes passengers more mobile. This is especially prominent in areas with more than one nearby airport. Many low-cost airlines have strategically negotiated lower-than-usual airport fees by threatening to take planes and passengers elsewhere (Dobruszkes, 2006) and there is evidence that passengers will travel considerable distances over land for lower airfares (Fuellhart, 2003, 2007; Matisziw and Grubesic 2010, 2011). For example, the presence of Southwest at BWI has been shown to draw a significant number of passengers from Harrisburg International (Fuellhart, 2003, 2007). The presence of Hapag-Lloyd airline at Hannover is thought to be the reason that this regional airport draws significant numbers of passengers from major metropolitan regions across North and West Germany (Pantazis and Liefner, 2006). This can lead to a functional differentiation of airports in multi-airport regions; with large airports serving international traffic and hub/transfer passengers, and smaller airports serving regional and low-cost airfares. However some airports have become so congested from direct traffic that they cannot accommodate significant hub traffic, such as Oakland and La Guardia (Derudder et al. 2010).

Geography of Airport Delays

Airport delays are growing problem for airlines and passengers in the United States.

Delays cost airlines at least $176 million a month (Ferguson, et al. 2010) and the cost to the U.S. economy has been estimated to be between $32.9 billion (Nextor et al., 2010) and $41 billion (Schumer, 2008) per year. An additional minute of delay was found to cost the airline up to $2.44 per passenger in a study based on a dramatic sudden increase in delays at La Guardia (Forbes, 2008). In 2000 the FAA spent $860 million to address delay problems and while this spending seems to have been effective, it remains questionable how much further these air-traffic control related improvements can continue to help effectively (Morrison and Winston, 2008). The problem is a complicated one and many different factors are involved. Some, such as weather, are largely unpredictable and uncontrollable and most literature focuses on the (seemingly) more manageable problem of congestion-related delays. Iterative Nash-equilibrium convergence has been used to show that even with increased capacity, delays will continue to grow throughout the country, reaching as much as an hour on average at O'Hare (Evans-Schäfer, 2011). Other simulations show that delays mostly occur at the airports affected by adverse conditions, but tend to propagate and accumulate at airports with strong capacity constraints (Pyrites et. al. 2012).

Airport congestion occurs when more runway slots or air-traffic control is demanded at an airport than is available. Generally it occurs at busy airports in regions with high population, hubs or airports which are unable to grow to meet demand (Santos and Robin, 2010). Planes that are unable to land immediately are generally required to orbit the airport until a runway slot becomes available, which wastes not only time, but huge quantities of fuel (Hansen, 2002). Occasionally aircraft will be delayed before take-off in anticipation of a congestion delay. With the exception of Washington National and

La Guardia, most U.S. airports are required to land aircraft on a first-come first-served basis, and this can result in significant delays, especially where different-size aircraft mix (Hansen, 2002).

The size of aircraft does impact congestion (it takes slightly longer to land a large aircraft) but the total number of aircraft has a much larger impact (Hansen, 2002). Unfortunately, airport landing fees are generally based on aircraft weight, creating an incentive to use larger numbers of smaller aircraft (Flores-Fillol, 2010). Rather than increase the size of aircraft in order to alleviate congestion there has been a slight trend toward smaller aircraft (Swan, 2002). Airlines are strongly cost-constrained by a competitive market where most companies operate below their margins and have reached the limit of being able to compete on fare prices and are being forced to compete by offering more convenient scheduling (i.e. more frequent flights, Brueckner, 2004). Airlines employing hub-and-spoke network models also have significant incentives to increase traffic at their hubs as each flight there effectively serves a much greater market and to schedule flights at peak-traffic times in order to decrease the length of layovers (Mayer and Sinai, 2003).

Alleviating Airport Congestion Delays without Increasing Capacity

Given the seriousness and the continued worsening of airport congestion and the significant challenges to airport capacity expansion, a number of interventions have been proposed to at least mitigate—if not solve—the problem. There are two dominant models, both centered on reducing the total number of flights (Debbage, 2002; Zografos, 2008). The first is slot-allocation which is already very common in Europe, but not used in the U.S. outside of the two most congested airports (La Guardia and Washington

National). Essentially, a set number of take-offs and landings, during specific time frames (usually the hour of the day) are allowed, and all other takeoffs and non-emergency landings are forbidden. These slots are then allocated based on different schemes. They are generally traded among airlines in an open market thereafter. The second system is congestion pricing where airlines are charged a fee for each flight based on how much congestion it causes, in order to discourage congesting flights and to offset the costs imposed on other airlines and passengers. There are a number of proponents for both systems. Airlines tend to favor slot-allocation because rather than being charged additional fee, they instead gain ownership of a new commodity, but dominant airlines stand to gain more than smaller competitors if slots are grandfathered in(Debbage 2002; Zografos 2008).

Slot-allocation systems are attractive to regulators because they explicitly limit the number of flights to a level within capacity, but they have several problems. The ideal is for slots to be traded on an open market, but this is seldom truly the case (Debbage, 2002). The systems in place in the New York-London market functionally limit inter-airline transfers which decreases competition and increases prices (Debbage, 2002). Also when slots are initially allocated, the ability of new airlines to enter into a market is limited, further decreasing competition. Some evidence exists that dominant airlines may hoard grandfathered slots in order to limit competition from other airlines (Debbage, 2002). This requires some kind of periodic redistribution of slots, either through trade, or through forcible seizure and auction, which may be interpreted as a legal taking and disallowed (Debbage, 2002).

Congestion-pricing has more proponents in academia and may well be considered more acceptable to most industry participants (Madas and Zografos, 2008). However, there is lively debate concerning how to best implement it. If the price is too low, congestion will continue to grow, but if it is too high, airlines and passengers will both suffer from higher fares and reduced services (Brueckner and Van Dender, 2008). Calculating the total congestion cost of any given flight is not a trivial matter, but much of the debate surrounds whether, or to what extent, airlines internalize self-imposed delays. When an airline schedules a flight to a congested airport, it inevitably affects its other flights at that airport to some extent. At a hub airport, where one airline dominates and a large percentage of that airline's total traffic is routed this effect could become very widespread. Thus if the airline were to increase its own traffic, it would impose delays primarily on itself (Debbage 2002; Flores-Fillol 2010). One would imagine that it will take these costs into account and that they will already be accounted for in its ticket prices. There are a number of economic-theoretical arguments concerning whether are or not this will be the case. Initially it was presupposed not only that airlines would internalize these costs, but that due to scheduling patterns, they would impose almost no delay on other airlines by doing so (Mayers and Sinai, 2003). However recent evidence shows that hub airlines frequently tend to create scheduling bottlenecks without regard to the delays they cause themselves or their competitors (Daniel and Hardback, 2009). No difference in delay propagation has been found between hub and non-hub airports (Diana, 2009) although some difference in local delays exists. The supposed effects of delay internalization are probably offset by competition with other airlines. For instance, when hub airlines voluntarily reduced peak-traffic at O'Hare, competitors quickly moved flights into those spots, resulting in no reduction in congestion and a net-loss for the hub airlines

(Daniel and Harback, 2008, 2009).

A number of models have been used to show that congestion-pricing would reduce congestion by reducing the number of flights and moving flight times away from peak-traffic times ('smoothing' the schedule). A stochastic bottleneck model suggests that at hub airports congestion pricing would primarily have the effect of spreading flight times out during each hour of the day while early and later departures would become more likely at all airports (Daniel and Harback, 2009). However, this fails to account for the decreased utility that this could cause by limiting passengers' access to convenient midday departures,easy-to-catch connections and shorter layovers. It has also been suggested that congestion-pricing would provide an incentive to airlines to increase aircraft size and decrease flight frequency (Hansen, 2002; Flores-Fillol, 2010).

While there is a wealth of theoretical research concerning the connection between congestion and airline scheduling behavior, empirical studies are limited. One empirical study of flight frequencies indicated that airlines tend to serve delay-prone airports with both smaller planes and less frequent flights (Pai, 2010). It showed that airlines appear to offer more frequent flights on competitive routes and to cater to managers and other wealthy populations with higher frequency flights. However this study examined only one year (2005) and treated early arrivals and late arrivals equivalently despite the fact that early arrivals are actually a boon to passengers and are less likely to contribute to congestion. By examining the relationship between delays and flight frequencies/aircraft size over a period of several years it should be possible to understand whether airlines respond to delays by altering their schedules or whether the frequency differences were inherent in the airports themselves.

Research Methods

Question

To what extent do U.S. Airlines alter their routes and schedules in response to airport based delays?

Hypothesis

In order to decrease flight delays at airports with high levels of delay, airlines should increase seats per flight, decrease available seats, decrease flights per day and/or schedule peaking at airports with higher delays. Due to economic pressures and recent trends the actual implementation of these policies are not expected to be widespread. An airline may be more likely to adopt delay mitigation strategies at an airport where it serves a greater proportion of the overall traffic because it will wield more control and will reap more of the benefits of delay decreases.

Data

I relied primarily on secondary data from the Bureau of Transportation Statistics. The vast majority of the data comes from the BTS's Airline On-Time Performance database which records airline, tail-number, origin, destination, scheduled departure, actual departure, and a breakdown of the types delays present for delayed flights for US air carriers that account for at least 1% of domestic traffic on a daily basis. These data are currently available for 1987-2011. However, the detailed breakdown of delays is currently only available for 2006-2011. I therefore have to limit my study to this time period. During this time period I have records for 40,592,740 flights, or roughly 6.8 million flights per year.

The major limitations of these data are that they do not explicitly include one variable of interest to this project. Seats per flight have a significant influence on congestion. Fortunately the number of seats on a given aircraft are recorded in the FAA's financial reporting from airlines. The number of seats as well as the tailnumber are recorded in these reports which are tabulated and available for download. Theoretically, these tables should have seat information for every aircraft flown by the airlines in the On-Time Performance database for the given time period. In practice there are many tailnumbers recorded in the on-time data that are not present in the aircraft ownership reports and vice-versa. 26.8% of the total number of flights are affected by this issue.

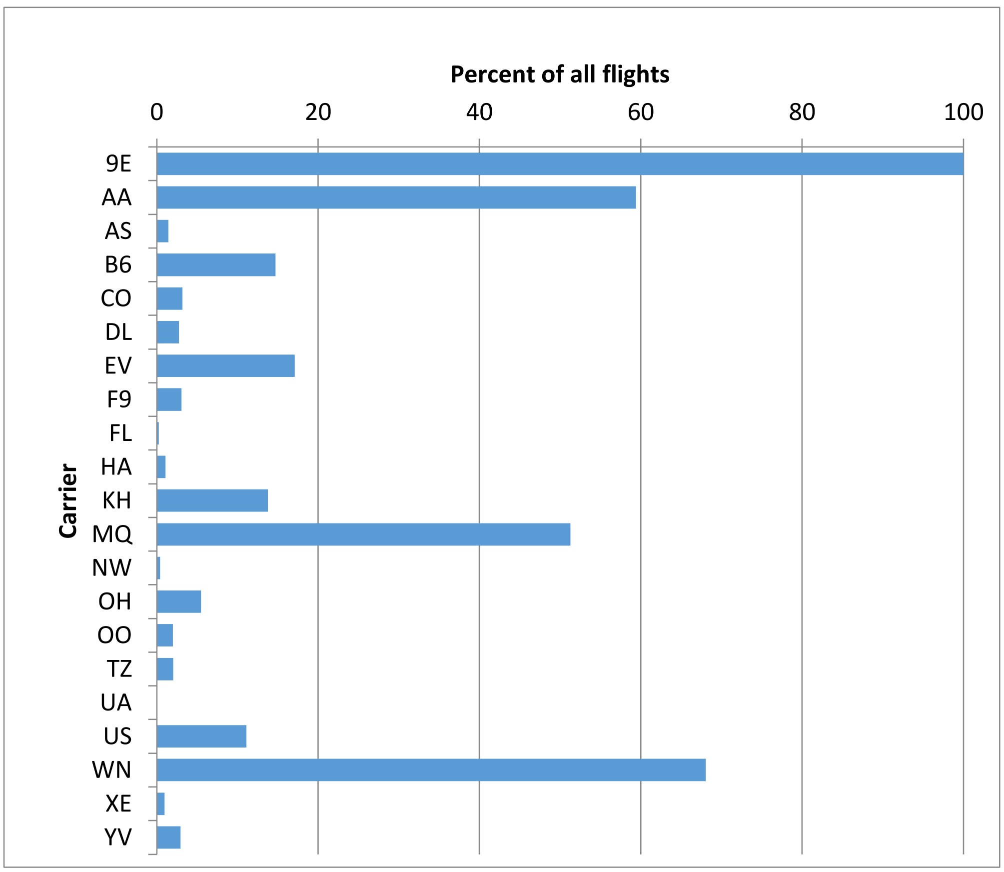


Figure 1 Mismatched Tailnumbers by Airline

Unfortunately, this proportion varies widely across airlines and airports. The fact that both of these data sets are self-reported by the airlines and that there is such variation between airlines (see

Figure 1) suggests that the discrepancies are caused by some part of the airlines' respective paperwork management policies. Most likely these mismatches arise as typographic errors. The variation between airports suggests that some locations have better data management and checking. Due to the fact that a mistyped tailnumber is still generally a valid tailnumber, it is very difficult to resolve these issues. A researcher can resolve some of them by cross-checking tables and determining that, for instance many American Airlines flights are recorded with tailnumbers such as N123 when in fact they should most likely be N123AA. However, it is not possible to prove that this is accurate. Therefore no attempt will be made to determine seat number for flights without a matching tailnumber in the aircraft database. Flights without seat data will be excluded from the calculation of average seats per flight. Destinations where no seat data are available are excluded from the remainder of the analysis, even the portions that do not explicitly include seat information. Although there is some evidence of clustering and spatial auto-correlation in the mismatching of tailnumbers, it is difficult to separate this from the actual clustering of airports themselves, the populations they serve and regional preference among airlines. No significant correlation was found between the total traffic at an airport and the number of mismatches (see Table 1below)

Table 1 Correlation test for airport size (total flights) vs. mismatch percentage

Correlation	0.012687
Count	333
t	0.230843
p-value	0.59121

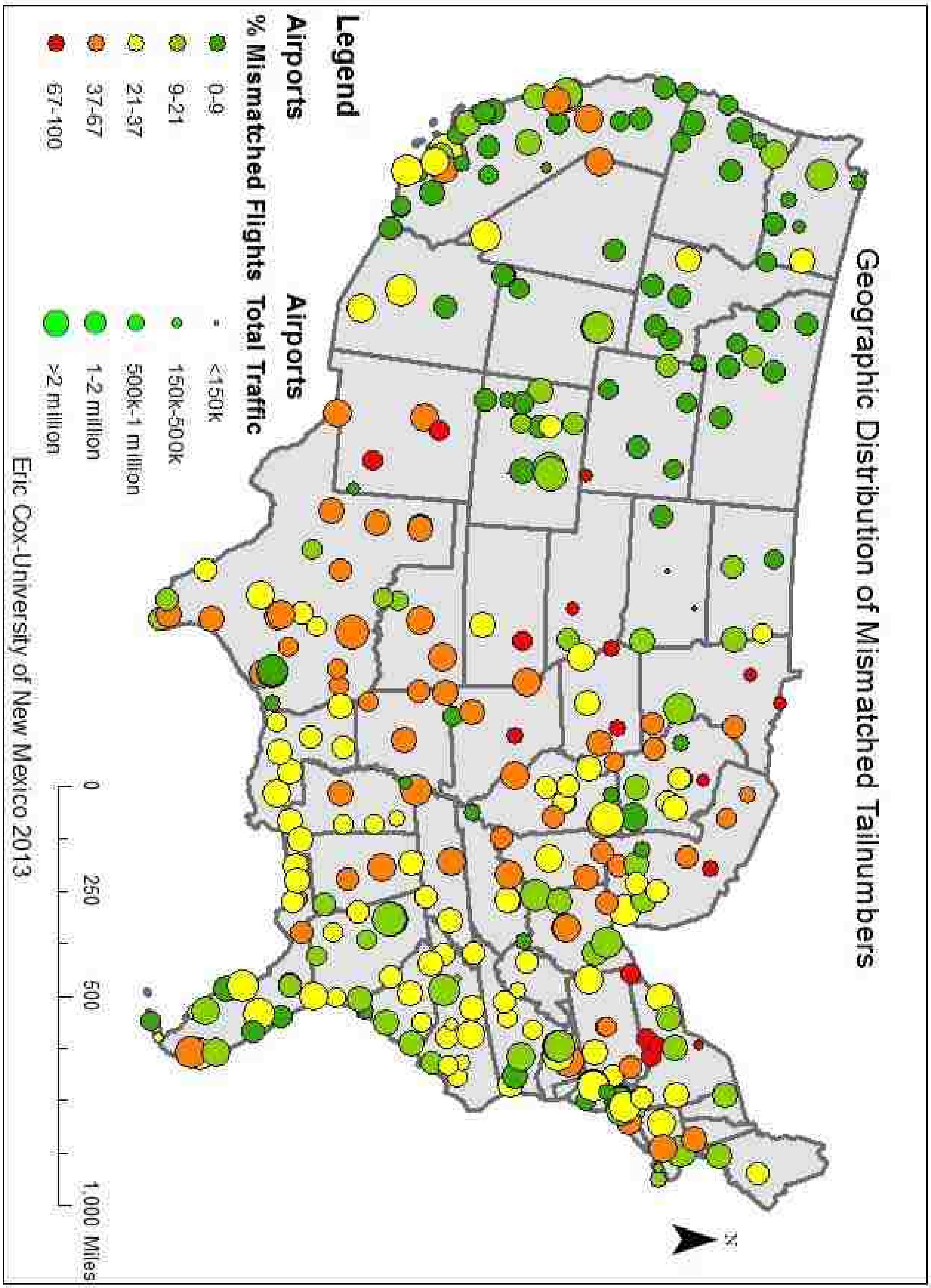


Figure 2 Mismatched tailnumbers by Airport

The fact that only carriers which account for at least 1% of traffic are recorded also presents a limitation in that it does not provide a complete picture of the domestic air traffic for the nation. However, small carriers by definition have a limited number of flights and therefore a smaller impact on congestion than large carriers. It is possible that a local carrier with a large number of very small flights could have a significant impact on congestion at a given airport but I think this effect can safely be ignored for two reasons. Firstly, I think the scenario is fairly unlikely. Secondly, any systemic delays caused by the flights of smaller carriers will still be reflected in the delay times recorded for larger carriers. The airlines included in the data set vary enough in size, primary airports and network strategy that they should provide a good picture of how airlines in general respond to airport-based delays. Based on the assumption that the smallest carrier in the dataset for a given year represents 1% (or more) of the total traffic we can estimate that the dataset includes at least 85-95% of the total traffic for the nation in any given year.

Data Preparation:

Before any analysis could be carried out the data had to be obtained, sorted, processed, combined, and summarized. The data were initially available as a series of text files (comma separated values) which I downloaded from the BTS/RITA site. In order to conduct the matching and summarizing steps the data needed to be converted into a more manageable format. I chose to use a relational database (PostgreSQL). Data were imported from their respective files (flights are broken up into monthly datasets) using a program that I wrote for the purpose (see Appendix A). This aspect of processing required approximately 120 hours of runtime due to the data volume.

Once the data were successfully imported into the database it was necessary to match the flight data with the seat number data from the aircraft ownership database. This involved a number of intermediate steps in order to allow the task of matching almost 41 million flights each with one of 15,000 aircraft to complete without the use of a supercomputer. The final matching was accomplished using the join procedure indicated in Appendix B. The delay variable of interest for each flight was calculated as the sum of the carrier delay (when the carrier holds a flight for some reason, usually passengers late from another flight or maintenance issues), national air system delay (imposed by the air traffic controllers) and late aircraft delay (flight did not depart on time due to a previous flight segment being delayed). This does not include delays coded as weather delays (which cannot be scheduled away) or as security delays (which are not related to air traffic). Although air traffic control delays are imposed on the airline by an external agent, they are likely due to congestion at the airport and are therefore included in the analysis.

Performing an analysis directly on the 1.4 billion data fields in the combined data set (40.6 million rows x 35 columns) was not a practical option and therefore the data had to be summarized. I choose to summarize the flight information based on the airline, the airport, and the month. It is important to retain the airline information because there is very strong evidence that different airlines employ different scheduling and routing tactics (Dobruszkes, 2006; Graham et al. 2006; Derudder and Witlox, 2009; Pai, 2010). Airport data are collected for both arrivals and departures at every airport in the dataset. Delays are broken up as departure delays and arrival delays in the original dataset.

Delays are also categorized for certain flights. Weather and security delays are subtracted from the departure delay and the departure delay is then subtracted from the arrival delay because a flight cannot be expected to arrive on time if it departs late. It is possible that favorable winds may allow it do so, but it is just as likely that it will be even further delayed en-route, therefore the expected time of arrival would be later. Data are summarized by month because this time period captures seasonal variations in air travel, but should not be affected by day-to-day or weekly variation. Thus every row of the summary table includes data aggregated for all flights arriving or departing at a specific airport for a specific carrier in a specific month. The calculation of this information was performed using the summary procedure listed in Appendix B. A count is maintained of the number of flights, and the number of flights with seat data in order to establish whether sufficient seat data are available for the carrier/airport/month. Flights without seat data are not included in the calculation of the average number of seats per flight. Seats and delay data are calculated separately for arrivals and departures and a weighted average of seats per flight and total delay is then determined based on the total number of arrivals and departures for the month. The number arrivals and departures are not always precisely equal over the course of month (although they are very close). This is due to a number of factors such as flights arriving after midnight on last day of the month.

In order to measure schedule peaking, the time between flights for each airline/airport is measured and the standard deviation normalized by dividing by the average time. This provides a measure of how closely flights are scheduled together that does not depend on the number of flights per day. Lowering or raising the number of flights should not change this measure so long as a consistent spacing is maintained, whereas moving flight times closer together or farther apart will affect it. An estimate of the number of seats offered by each airline at each airport for each month ('available seats') was produced. While seats per flight and flights per month both measure aspects of the total traffic to an airport, they do not measure it directly. Either one could change as a result of altering the fleet used to serve the airport without affecting the number of passengers that the airline could physically convey to the airport. Available seats are estimated by multiplying the total number of flights by the average number of seats per flight. As such, it is only as good as our seats estimate. If this measure decreases, it should indicate that the airline is reducing service to the airport in question, but it is not easy to determine whether it is because demand has decreased, the airline is no longer competitive in that market, or because the airline is rerouting transit passengers. Airline dominance for an airport is estimated by the proportion of total flights that the airline operates at that airport. It therefore ranges from 0 to 1 (although it would not be calculated where it is exactly zero).

Airline/airport pairs which did not have data for the full 72-month study period were excluded from the final analysis. Airline/airport pairs which did not have seat data were also excluded from further analysis so that all tests would be run on the same set of data. Some airline/airport pairs were served by so few flights that no meaningful measure of peaking could be developed. These were also excluded. This step left 798 pairs out of 1742 original pairs. This is due to a large number of pairs where service was either stopped or started during the study period. By and large this represented either the termination or the beginning or very infrequent services, once daily or once weekly. Such infrequent services are inherently tenuous links to the airline's route map as a whole and can quickly be dropped when they do not earn enough or are quickly started when an opportunity presents itself.

Analysis

The amount of data available for this project is staggering, but at the same time the number of factors that affect airlines' routes and schedules is also very large. The analysis is designed to use the data to ferret out what may be small differences within several transcontinental networks over a period of years. By only comparing like with like I hope to exclude much of the variation for which I cannot account. For that reason I examined changes based on the same airline at the same airport in the same month of the year. This should hold constant airline effects, such as different priorities, routing strategies, and established operations. It should also eliminate most seasonal effects (holiday traffic, etc.) to the extent that they remain constant over the study period. Finally it should eliminate airport effects such as number or runways, runway configuration, gate assignment and prevailing winds. I have also, where possible, ignored delays that are not the airlines' fault and cannot be predicted, such as weather and security delays.

The goal of the analysis is to determine where changes in flights, seats, or schedule peaking are related to the level of delay, whether these airports/airlines are affected by or are effected in response to a higher level of delay, and finally whether airlines focus on airports where they are more dominant. In order to answer this question a two-step analysis is performed. The first step is multiple regression analyses (not multiple regression analysis) to determine where there is a significant relationship between delay and variables of interest. The model used is $v_{interest} = \beta_0 + \beta_1 \cdot average\ delay$ where $v_{interest}$ is seats per flight, total flights, peaking or available seats. The second step is a comparison to determine whether or not there are differences in delay between airports/airlines with a significant relationship and those without and another comparison to determine whether delay mitigation strategies are more common at airports where the airline being examined is dominant.

The regression analysis is performed separately for each airport/airline pair. The changes in seats per flight, total flights, schedule peaking and available seats are determined based on the same months in subsequent years in order to minimize the effects of seasonal variations in traffic. A sample regression is seen in Figure 3 below.

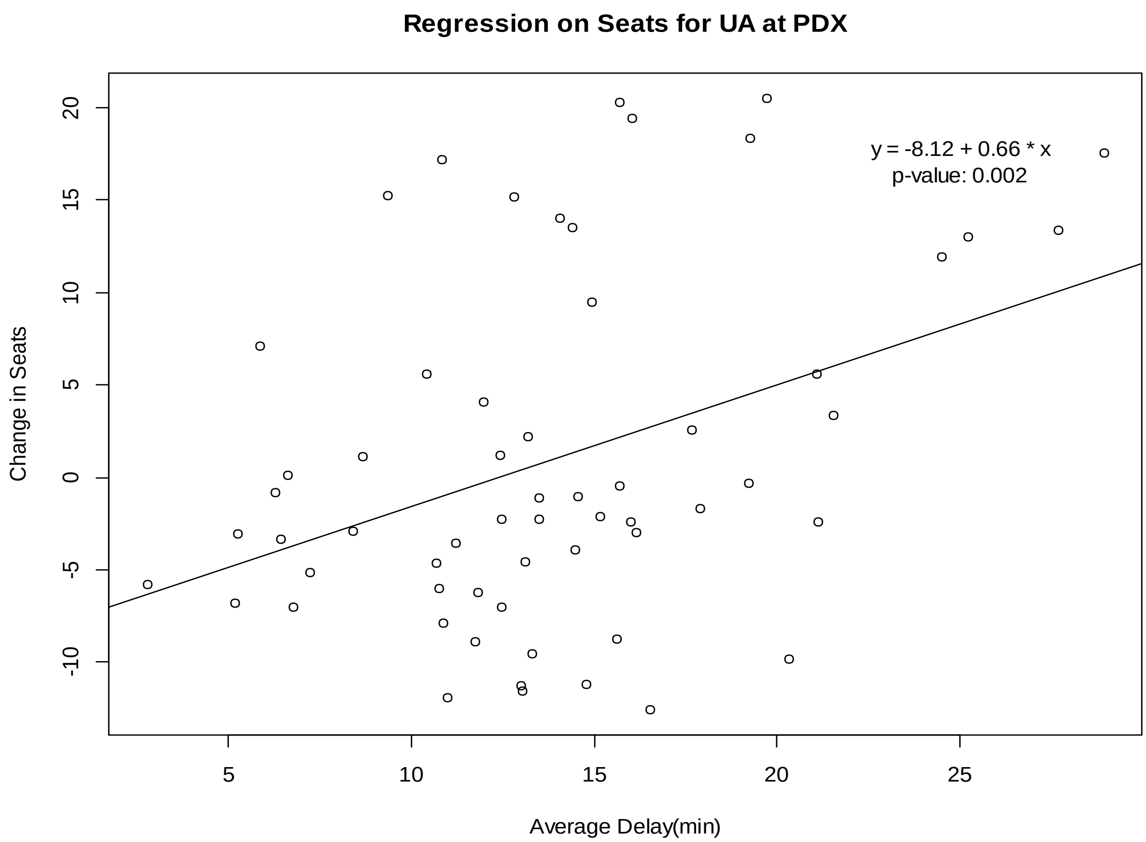


Figure 3 Sample regression: Seats per Flight for United at Portland (study group)

Due to the fact that airlines need to announce flight schedules 6-12 months in advance it is unlikely that we could see scheduling responses on a month-to-month basis. Due to the

6 year study period this leaves us with 5 year-to-year intervals or a total of 60 monthly periods for each airline/airport pair. Each of these three change variables is then subjected to a separate regression analysis using the total delay levels as the explanatory variables. The significance level and the slope coefficient for each regression analysis are recorded for each airline/airport pair.

Based on the results of the regression analysis airlines and airports are separated into study and control groups for each variable. Those airlines and airports that show a significant ($\alpha=.05$) relationship between delay changes in scheduling variables and a slope with the expected sign (positive for seats per flight and negative otherwise) are included in the study group and all other airlines and airports are included in the control group. Due to apparent differences and non-normality in the distributions of the study groups Welch's T-Test is used to determine whether the two groups show a significant ($\alpha=.05$) difference in average delay for each of the three variables of interest. It is also used to determine whether there is a significant difference in airline dominance between the study and control groups. The study group for seats per flight was found to have 113 members. The total flights group had 158 members. The schedule peaking group had 145 members. The available seats group had 153 members. The control group in each case represented the remainder of the 798 full test cases. The study groups for each variable are listed in full in Appendix F.

Results

Delay

The full results of the t-tests for delay can be seen in Table 2. Based on these results we can conclude that the level of delays for airline/airport pairs that show a significant relationship between delay and changes in schedule peaking are higher than the level of delays at airline/airport pairs where they do not. No significant difference was observed for the other variables. Schedule peaking is arguably the simplest change to implement in that it does not require any change in the routes or equipment.

Variable	p-value	Control mean (minutes delay)	Study mean (minutes delay)	t	df
Seats per flight	.6513	15.37	15.11	.454	68.8
Flights	.8086	15.34	15.49	-.243	108.52
Peaking	.03153	15.21	16.83	-2.78	76.59
Available Seats	.5382	15.32	15.68	-.705	108.43

Table 2 T-Test Results for Delay by Study Group

The observed difference in the means for peaking is fairly small (less than two minutes). Examining the distribution of total delay for the peaking study group shows a much thicker right tail (see Figure 4). Specifically, it shows a 'knee' around 20 minutes of delay where we see many more flights above that level than expected. This suggests that 20 minutes may serve as a threshold to determine when delays have become unacceptable and need to be addressed by an intervention. Another probable explanation for why the differences in the means are so low is that airlines (we can assume) manage their schedules differently and while some airlines may not implement a specific intervention to say reduce schedule peaking based on delays, those that do will be lumped together with those that do not.

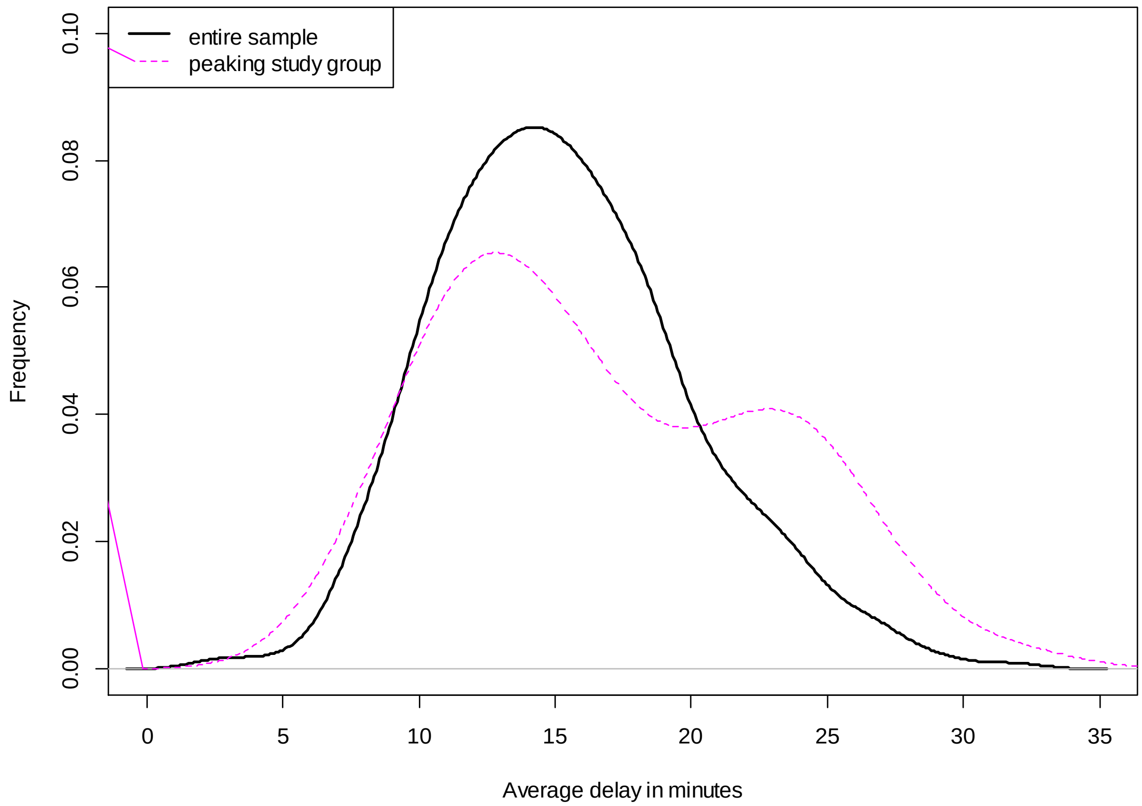


Figure 4 Delay distribution for entire sample and peaking study group.

The regressions determined where delays drive schedule changes. The first tests determined that there are limited global trends. Next we need to determine whether specific airlines are adjusing schedules strategically. To determine where there are airline effects the same t-test was implemented on an airline-by-airline basis to determine which airlines are implementing which types of interventions. The results are shown below. A Bonferroni correction has been applied resulting in an adjusted α=.05/44=.001136. Peaking for Continental (highlighted in yellow below) was the only test with a p-value under .05. However, after the Bonferroni Correction no significant results were found for any airline or any variable.

Table 3 T-Test Results for Delay by Airline and Study Group
Airline/variable pairs where the study group had two or fewer members are excluded.

Airline	Seats/flight	Total flights	Peaking	Available Seats
AA	0.905030033	0.058329156	0.181148118	0.079612185
AS	0.890515108	0.99810044	0.658741871	0.99810044
CO	0.996871793	0.097984686	0.04653525	0.097984686
DL	0.385622199	0.628344371	0.141562062	0.628344371
EV	0.611999659	na	0.623961072	na
F9	na	0.227120341	na	0.227120341
FL	0.365627236	0.511897117	na	na
HA	na	0.620079618	na	0.620079618
OO	0.751023051	0.71296629	0.754394193	0.823951608
UA	0.205219739	0.740546164	0.55126813	0.704809012
US	0.475332768	0.313039416	0.195626184	0.025926768
WN	na	na	0.755542194	na
XE	0.265251541	0.174550184	0.089210118	0.174550184
YV	0.847592101	0.687166815	na	0.447401696

Had a parametric test been used instead of Welch's Test we would not have had to exclude as many pairs and some of the excluded airline/variable pairs would have yielded significant results, but the evidence does not support the assumption of heteroscedasticity or normality.

Dominance

The full results of Welch's t-tests for dominance can be seen in Table 4. Based on these results we can conclude at a .05 level of significance that the level of airline dominance[1] for airline/airport pairs that show a significant relationship between delay and changes in seats per flight is lower than the level of dominance at airline/airport pairs where they do

[1] The proportion of total flights that an airline operates at an airport.

not. No significant difference was observed for the other variables. This result is somewhat counterintuitive given that the expected result was to have higher average dominance in the study group than in the control. It may not be related to delay-management per-se and may instead reflect airlines adopting a strategy based on competition with more dominant airlines.

Table 4 Results of Dominance T-Tests for the entire sample

Variable	p-value	Control mean (dominance)	Study mean (dominance)	t	df
Seats per flight	0.008801	20.61 %	13.58%	2.6886	76.459
Flights	0.02224	20.77%	15.00%	2.3143	128.25
Peaking	0.5816	20.25%	18.46%	0.5533	82.94
Available Seats	0.01534	20.79%	14.81%	2.4565	130.67

The distributions of the study group and the entire dataset show very similar curves except at a dominance level of 100% where an airline has a monopoly at an airport

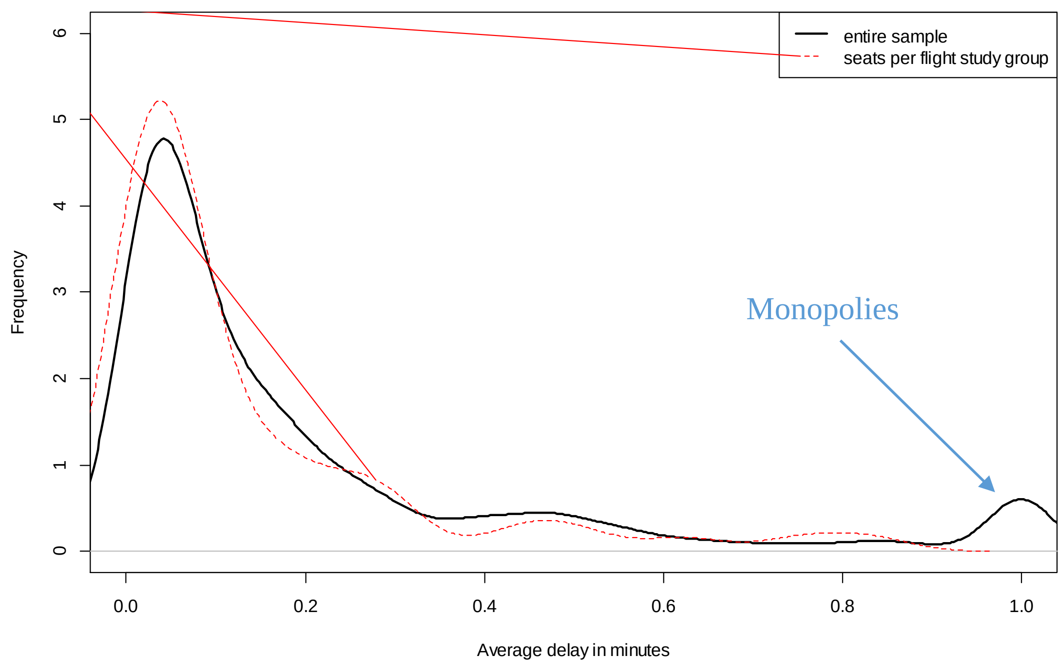


Figure 5 Airline Dominance Distribution

(see Figure 5). There are a number of monopoly destinations in the group as a whole, but none in the study group. If monopoly destinations are excluded from the test, the p-value jumps to .4—well above significance. There are probably a number of reasons why monopoly destinations don't show any seats-per-flight response. Where an airline has a monopoly, it has very little incentive to make changes other than to reduce costs. Monopoly destinations also tend to be much smaller airports where other airlines have little incentive to compete, and they are served by smaller 'regional' carriers (Alaska, American Eagle, ExpressJet, Comair, SkyWest and Mesa). These airlines are not likely to change their seats per flight for a number of reasons. With the exception of Alaska, none of them operate an aircraft with more than 90 seats. They simply do not have the option to make a large change in the number of seats they offer on a flight. These airlines mainly operate connecting flights from smaller destinations that nationwide airlines don't serve. As they operate at a large number of small airports, changes made at any one airport don't have much impact. In order to increase seats per flight at one of their connection airports they would need to increase it at all of the airports that feed into it.

Based on these factors I cannot conclude that airlines are significantly more likely to adopt any delay mitigation strategy at airports where they serve a relatively higher percentage of total traffic. Furthermore, we can conclude that they are very unlikely to do so at an airport where they have no competitors.

Conclusion

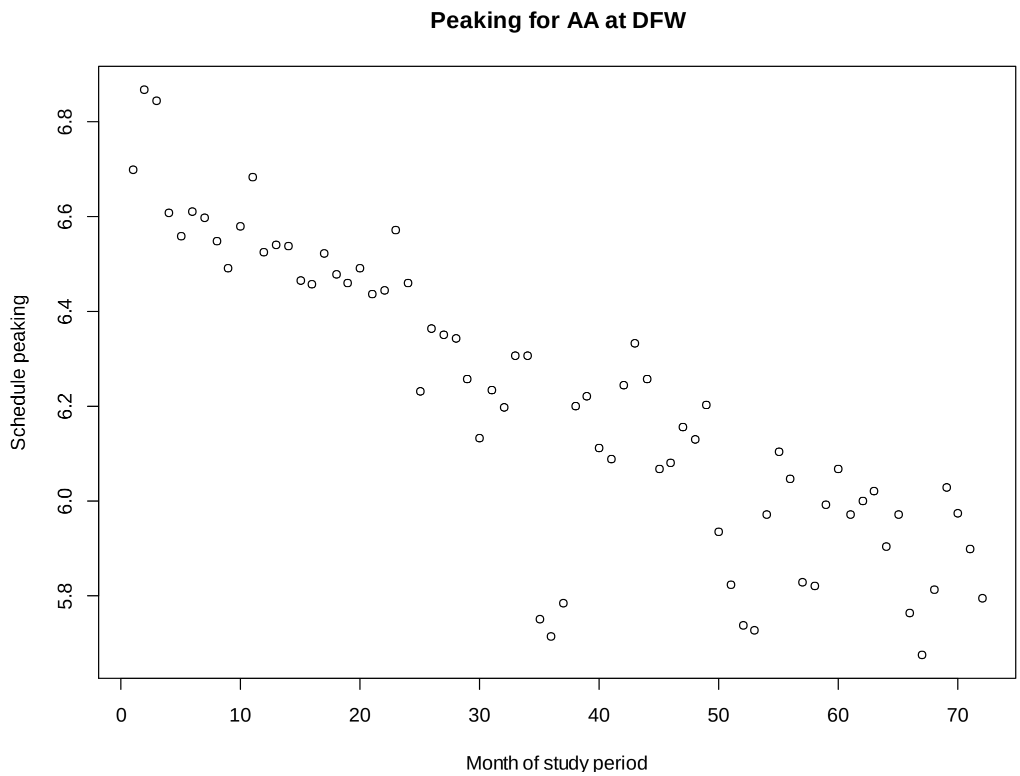
Overall the results of the study support the hypothesis that airlines do not respond to high levels of delay with the delay mitigation techniques that were studied. The exception is that there appears to be some evidence that airlines in general decrease schedule peaking at airports with higher levels of delay. However, no evidence was found of any particular airline systematically engaging in this strategy. This outcome may suggest that these changes are only happening at a limited number of the airports with the highest delay and that there are not enough from any single airline to show a significant trend. In fact a number of airlines had to be excluded from the individual tests because they had only one airport in the study group.

The secondary hypothesis that airlines would implement delay mitigation strategies at airports where they operated a higher proportion of the total number of flights was not borne out. There is good evidence that airlines will not respond to delays with seat changes at airports where they are the only carrier, but there was no evidence of any difference for airports with multiple carriers.

All attempts were made to use analysis techniques that control as many confounding factors as was feasible but in this field true controls are not possible. It is still certainly conceivable that airlines are implementing delay mitigation strategies that were not detected for one reason or another. Perhaps they use strategies other than those that were studied, such as improving the efficiency of departure and arrival procedures. A number of potentially influential factors were not taken into account in this study due to a lack of data or time constraints or the inherent difficulty of accounting for them. Although the study period is relatively short, a number of structural changes in the air travel network occurred during this time. A weak economy saw ticket sales drop and consequently there was a decline in air travel across the board, bucking the long term trend of increasing air travel. During the study period Northwest, which had been one of the larger airlines, was absorbed by Delta Airlines. It is difficult to determine how this should be accounted for given that the two were—at the beginning of the study period—real competitors but by the end of the study period all of Northwest's flights, aircraft and destinations had been taken over by Delta. This results in an apparent increase in total flights, available seats, etc. for Delta at certain airports. Because Northwest had ceased to exist by the end of the study period, it was automatically excluded from the final steps of the analysis.

Another potential issue is cases where strategies similar to delay mitigation are adopted, but not in apparent response to delays. For example, American Airlines has consistently been decreasing schedule peaking at DFW during the study period regardless of the level of delay in a given month (see Figure 6 on the next page). This does not appear to have been affected by the level of delay at DFW for AA and DFW does not have particularly high levels of delay compared to other AA destinations[2]. However, it can be argued that the policy may have prevented increased delay. As a key hub in American's network they may want to maintain low delays there rather than decrease them at smaller airports. This type of scenario is very difficult to account for.

[2] The level of delay at DFW is .1 standard deviations above the mean.



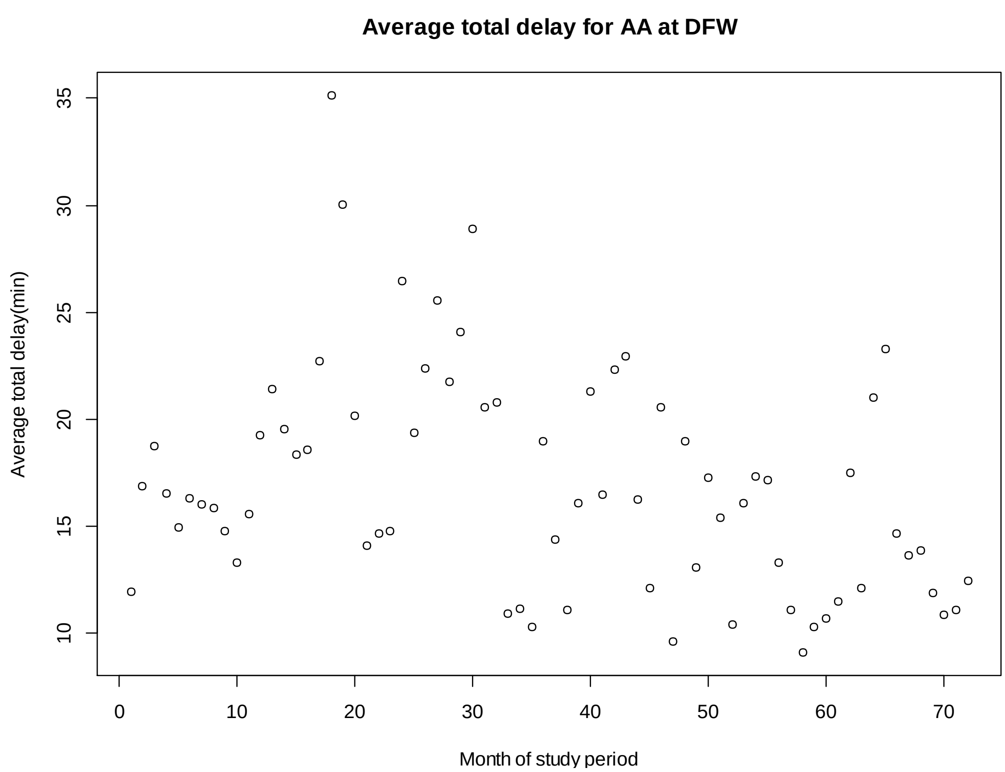


Figure 6 Peaking and Delay for AA at DFW

There is still substantial room to analyze these data further in order to address the question at hand. It would be useful to examine it with respect to network connectivity, examining routes directly. It would also be worthwhile to pursue in-depth studies of one or more individual airlines to determine what other factors may be in play. A longer study period would also be desirable and will become more practical as more data are available each month. Combining these analyses with interviews with airline personnel and reviews of airline documents would also be interesting, assuming one could find a cooperative airline to work with.

The question of whether airlines should be adopting these delay mitigation strategies is an open one. Every change to routes or schedules ultimately causes an inconvenience to many passengers, often to the benefit of others. Determining which tradeoffs are worthwhile is a difficult decision. However, reducing delay times, especially for landing flights does reduce the fuel consumption, and therefore the emissions and the cost of flights to an airport. Based on this research I would conclude that if we wish to reduce delays, these techniques still represent promising choices, but that it appears that the incentive structure that currently exists does not encourage airlines to use them. It may therefore be necessary to make a policy intervention necessary in order to decrease delays.

List of Appendices

Appendix A: C# Code written for the project

Aircraft Importation

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using System.Threading;
using System.IO;
using Npgsql;

namespace FlightsParser
{
    public partial class FlightForm : Form
    {
      Thread processingThread;
        System.Windows.Forms.Timer tmr;
        long bytesToRead, bytesRead;
        DateTime started;
        char[] delims = new char[] { ',' };
        delegate void TextBoxCallBack(TextBox tb, string s);
        delegate void ProgressBarCallBack();
        TextBoxCallBack tbcb;
        ProgressBarCallBack pgcb;
        Dictionary<string, string> routes;
        NpgsqlConnection dataConn;
        public FlightForm()
        {
            InitializeComponent();
        }

        private void button1_Click(object sender, EventArgs e)
        {
            if (folderBrowserDialog1.ShowDialog().Equals(DialogResult.OK))
            {
                textBox1.Text = folderBrowserDialog1.SelectedPath;
            }
        }
```

```
        private void button2_Click(object sender, EventArgs e)
        {
            try
            {
                dataConn = new
NpgsqlConnection("Server=localhost;Port=5432;DataBase=postgres;User
Id=postgres;Password=9a55w0rd");
                dataConn.Open();
            }
            catch (Exception ex)
            {
                MessageBox.Show("Cannot connect.\r\n" + ex.Message);
                return;
            }
            tbcb = new TextBoxCallBack(UpdateTextBox);
            pgcb = new ProgressBarCallBack(UpdateProgress);




            //flightList = new List<string>();

            if (MessageBox.Show("Clear existing data?", "",
MessageBoxButtons.YesNo).Equals(DialogResult.Yes))
            {
                NpgsqlCommand comm;
                comm = new NpgsqlCommand("DELETE FROM aircraft;", dataConn);
                comm.ExecuteNonQuery();
            }


            string[] files = Directory.GetFiles(textBox1.Text, "*.csv",
SearchOption.AllDirectories);

            processingThread = new Thread(ProcessFiles);

            processingThread.Start(files);
        }
        private void tmr_Tick(object sender, EventArgs e)
        {
            if (bytesRead >= bytesToRead)
            {
                ((System.Windows.Forms.Timer)sender).Stop();
```

```
            ((System.Windows.Forms.Timer)sender).Dispose();
            this.Invoke(tbcb, textBox9, "");
            return;
        }
        TimeSpan elapsed = DateTime.Now.Subtract(started);
        this.Invoke(tbcb, textBox8, elapsed.ToString());
        if (bytesRead > 0)
        {
            TimeSpan remaining = new TimeSpan(elapsed.Ticks / bytesRead * bytesToRead);
            remaining = remaining.Subtract(elapsed);
            this.Invoke(tbcb, textBox9, remaining.ToString());
        }
    }
    private void ProcessFiles(object obj)
    {
        string[] files = (string[])obj;
        for (int i = 0; i < files.Length; i++)
        {
            string file = files[i];
            this.Invoke(tbcb, textBox1,file);
            this.Invoke(tbcb,textBox5, String.Concat(i + 1, "/", files.Length));

            FileInfo info = new FileInfo(textBox1.Text);
            bytesToRead = info.Length;
            bytesRead = 0;

            if (tmr != null)
                tmr.Stop();
            else
                tmr = new System.Windows.Forms.Timer();
            tmr.Interval = 1000;
            tmr.Tick += new EventHandler(tmr_Tick);

            started = DateTime.Now;
            tmr.Start();
            Process(file);
        }
    }
    private void Process(object obj)
    {
        string fn = (string)obj;

        NpgsqlCommand comm, comm2;
```

```
            StreamReader sr = new StreamReader(fn);
            Dictionary<string,int> cols = new Dictionary<string,int>();
            string line = sr.ReadLine();
            string[] cls = line.Split(delims);
            bytesRead = line.Length + 1;
            for(int i=0; i<cls.Length; i++)
            {
                cols.Add(cls[i].Replace("\"",""),i);
            }
            int lines = 0, flights = 0, tails = 0;
            while (!sr.EndOfStream)
            {
                line = sr.ReadLine();
                lines++;
                bytesRead += line.Length+1;
                string[] toks = GetCSVFields(line);
                string y = NoQuotes(toks[cols["YEAR"]]);
                string tn = NoQuotes(toks[cols["TAIL_NUMBER"]]);
                string numSeats = NoQuotes(toks[cols["NUMBER_OF_SEATS"]]);
                string manu = NoQuotes(toks[cols["MANUFACTURER"]]);
                string mod = NoQuotes(toks[cols["MODEL"]]);
                string arlID = NoQuotes(toks[cols["AIRLINE_ID"]]);
                string uid = NoQuotes(toks[cols["UNIQUE_CARRIER"]]);
                string date = NoQuotes(toks[cols["ACQUISITION_DATE"]]);

                if (arlID == "")
                {
                    continue;
                }

    string query1 = "INSERT INTO aircraft
    (tailnumber,year,airline,seats,manufacturer,model,acq_date,airline_id) VALUES('";
    query1 += tn + "'," + y + ",'" + uid + "'," + numSeats + ",'" + manu + "','" + mod + "','" + date
    + "'," + arlID + ")";

    string query2 = "UPDATE flight SET seats = " + numSeats + " WHERE tailNumber = '" + tn +
    "' AND seats IS NULL";


                try
                {
```

```
            comm = new NpgsqlCommand(query1, dataConn);
            tails += comm.ExecuteNonQuery();

            //comm2 = new NpgsqlCommand(query2, dataConn);
            //comm2.CommandTimeout = 1000;
            //flights += comm2.ExecuteNonQuery();
        }
        catch (Exception ex)
        {
            MessageBox.Show(ex.Message);
        }

        this.Invoke(tbcb, textBox2, lines.ToString());
        //this.Invoke(tbcb, textBox3, flights.ToString());
        this.Invoke(tbcb, textBox7, y + "_" + uid + "_" + tn + "_");
        this.Invoke(tbcb, textBox4, tails.ToString());
        this.Invoke(pgcb);
    }
    sr.Close();
}
private static string[] GetCSVFields(string s)
{
    List<string> fields = new List<string>();
    string s2 = s.Substring(0);
    int sIndex = 0;
    int eIndex = 0;
    bool hasQuote = false;
    for (int i = 0; i < s2.Length; i++)
    {
        char c = s2[i];
        if (c == ',')
        {
            if (!hasQuote)
            {
                eIndex = i - 1;
                fields.Add(s2.Substring(sIndex, eIndex - sIndex + 1));
                sIndex = i + 1;
            }
        }
        if (c == '"')
        {
            hasQuote = !hasQuote;
        }
```

```
        }
        fields.Add(s2.Substring(sIndex, s2.Length - sIndex));
        return fields.ToArray();
    }
    private void UpdateTextBox(TextBox tb, string s)
    {
        tb.Text = s;
        tb.Refresh();
    }
    private void UpdateProgress()
    {
progressBar1.Value = Math.Min(progressBar1.Maximum,
Math.Max(progressBar1.Minimum, (int)((progressBar1.Maximum -
progressBar1.Minimum) * (bytesRead / (double)bytesToRead) +
progressBar1.Minimum)));
    }
    private static string NoQuotes(string s)
    {
        return s.Replace("\"", "");
    }
    private static string Elapsed(string s1, string s2)
    {
        s1 = NoQuotes(s1);
        s2 = NoQuotes(s2);
        string h1, m1, h2, m2;
        if (s1 == "" || s2 == "")
            return "0";
        if (s1.Contains(':'))
        {
            h1 = s1.Substring(0, s1.IndexOf(':'));
            m1 = s1.Substring(s1.IndexOf(':') + 1);
            h2 = s2.Substring(0, s1.IndexOf(':'));
            m2 = s2.Substring(s1.IndexOf(':') + 1);
        }
        else
        {
            h1 = s1.Substring(0, s1.Length - 2);
            h2 = s2.Substring(0, s2.Length - 2);
            m1 = s1.Substring(s1.Length - 2);
            m2 = s1.Substring(s2.Length - 2);
        }
        return "0";
        int hr1 = Convert.ToInt16(h1);
```

```
            int min1 = Convert.ToInt16(m1);
            int hr2 = Convert.ToInt16(h2);
            int min2 = Convert.ToInt16(m2);

            TimeSpan el1 = new TimeSpan(hr1, min1, 0);
            TimeSpan el2 = new TimeSpan(hr2, min2, 0);
            TimeSpan el = el2.Subtract(el1);
            if (el.TotalMinutes < 0)
                el.Add(new TimeSpan(24, 0, 0));
return String.Concat(el.TotalHours.ToString("00"), ":", el.TotalMinutes.ToString("00"));
        }
    }
}
```

Flight Importation

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using System.Threading;
using System.IO;
using Npgsql;

namespace TableParser
{
    public partial class Form1 : Form
    {
        Thread processingThread;
        System.Windows.Forms.Timer tmr;
        long bytesToRead, bytesRead;
        DateTime started;
        char[] delims = new char[] { ',' };
        delegate void TextBoxCallBack(TextBox tb, string s);
        delegate void ProgressBarCallBack();
        TextBoxCallBack tbcb;
        ProgressBarCallBack pgcb;
        List<string> airports, airlines;//, flightList;
        Dictionary<string, string> routes;
```

```
        NpgsqlConnection dataConn;
        string targetCarrier = "AA";
        public Form1()
        {
            InitializeComponent();
        }

        private void button1_Click(object sender, EventArgs e)
        {
            if(folderBrowserDialog1.ShowDialog().Equals(DialogResult.OK))
            {
                textBox1.Text = folderBrowserDialog1.SelectedPath;
            }
        }

        private void button2_Click(object sender, EventArgs e)
        {
            try
            {
dataConn = new NpgsqlConnection("Server=localhost;Port=5432;DataBase=postgres;User
Id=postgres;Password=9a55w0rd");
              dataConn.Open();
            }
            catch (Exception ex)
            {
                MessageBox.Show("Cannot connect.\r\n" + ex.Message);
                return;
            }
            tbcb = new TextBoxCallBack(UpdateTextBox);
            pgcb = new ProgressBarCallBack(UpdateProgress);

            targetCarrier = textBox10.Text;


            airlines = new List<string>();
            airports = new List<string>();
            //flightList = new List<string>();

if (MessageBox.Show("Clear existing data?", "",
MessageBoxButtons.YesNo).Equals(DialogResult.Yes))
            {
                NpgsqlCommand comm;
                comm = new NpgsqlCommand("DELETE FROM sw_flight;", dataConn);
```

```
            //comm.ExecuteNonQuery();
            //comm.CommandText = "DELETE FROM airline;";
            //comm.ExecuteNonQuery();

            comm.ExecuteNonQuery();
        }


/*NpgsqlDataAdapter da = new NpgsqlDataAdapter("SELECT airline_id, code FROM
airline;", dataConn);
        DataTable dt = new DataTable();
        da.Fill(dt);
        foreach (DataRow dr in dt.Select())
        {
            airlines.Add(dr["airline_id"].ToString());
        }
        dt = new DataTable();
da = new NpgsqlDataAdapter("SELECT airport_id, code FROM airport", dataConn);
        da.Fill(dt);
        foreach (DataRow dr in dt.Select())
        {
            airlines.Add(dr["airport_id"].ToString());
        }
        dt = new DataTable();
        da = new NpgsqlDataAdapter("SELECT flightid FROM flight", dataConn);*/
        //da.Fill(dt);
        /*foreach (DataRow dr in dt.Select())
        {
            flightList.Add(dr["flightid"].ToString());
        }*/


string[] files = Directory.GetFiles(textBox1.Text, "*.csv",SearchOption.AllDirectories);

        processingThread = new Thread(ProcessFiles);

        processingThread.Start(files);

    }
    private void tmr_Tick(object sender, EventArgs e)
    {
        if (bytesRead >= bytesToRead)
        {
```

```
        ((System.Windows.Forms.Timer)sender).Stop();
        ((System.Windows.Forms.Timer)sender).Dispose();
        this.Invoke(tbcb, textBox9, "");
        return;
    }
    TimeSpan elapsed = DateTime.Now.Subtract(started);
    this.Invoke(tbcb, textBox8, elapsed.ToString());
    if (bytesRead > 0)
    {
        TimeSpan remaining = new TimeSpan(elapsed.Ticks / bytesRead * d);
        remaining = remaining.Subtract(elapsed);
        this.Invoke(tbcb, textBox9, remaining.ToString());
    }
}
private void ProcessFiles(object obj)
{
    string[] files = (string[])obj;
    for (int i = 0; i < files.Length; i++)
    {
        string file = files[i];
        this.Invoke(tbcb, textBox1,file);
        this.Invoke(tbcb,textBox5, String.Concat(i + 1, "/", files.Length));

        FileInfo info = new FileInfo(textBox1.Text);
        bytesToRead = info.Length;
        bytesRead = 0;

        if (tmr != null)
            tmr.Stop();
        else
            tmr = new System.Windows.Forms.Timer();
        tmr.Interval = 1000;
        tmr.Tick += new EventHandler(tmr_Tick);

        started = DateTime.Now;
        tmr.Start();
        Process(file);
    }
}
private void Process(object obj)
{
    string fn = (string)obj;
```

```
            NpgsqlCommand comm;


            StreamReader sr = new StreamReader(fn);
            Dictionary<string,int> cols = new Dictionary<string,int>();
            string line = sr.ReadLine();
            string[] cls = line.Split(delims);
            bytesRead = line.Length + 1;
            for(int i=0; i<cls.Length; i++)
            {
                cols.Add(cls[i].Replace("\"",""),i);
            }
            int lines = 0, flights = 0;
            while (!sr.EndOfStream)
            {
                line = sr.ReadLine();
                lines++;
                bytesRead += line.Length+1;
                string[] toks = GetCSVFields(line);

                string fltNum = NoQuotes(toks[cols["FL_NUM"]]);
string dateString = toks[cols["YEAR"]] + toks[cols["MONTH"]].PadLeft(2, '0') +
toks[cols["DAY_OF_MONTH"]].PadLeft(2, '0');
                string cc = NoQuotes(toks[cols["UNIQUE_CARRIER"]]);
                if (cc != targetCarrier)
                    continue;
                string orig = NoQuotes(toks[cols["ORIGIN"]]);
                string dest = NoQuotes(toks[cols["DEST"]]);
string fltID = cc + fltNum.PadLeft(5, '0') + "_" + dateString + "_" + orig + "_" + dest;


                if (true)
                {


                    if (toks[cols["CRS_ELAPSED_TIME"]] == "")
                    {
toks[cols["CRS_ELAPSED_TIME"]] = Elapsed(toks[cols["CRS_DEP_TIME"]],
toks[cols["CRS_ARR_TIME"]]);
toks[cols["ACTUAL_ELAPSED_TIME"]] = Elapsed(toks[cols["DEP_TIME"]],
toks[cols["ARR_TIME"]]);
toks[cols["AIR_TIME"]] = Elapsed(toks[cols["WHEELS_OFF"]], toks[cols["WHEELS_ON"]]);
                    }
```

```
                string airlineId = NoQuotes(toks[cols["AIRLINE_ID"]]);
                /*if (!airlines.Contains(airlineId))
                {
                    airlines.Add(airlineId);
 comm = new NpgsqlCommand("INSERT INTO airline (code,airline_id) Values('" + cc + "',"
+ airlineId + ");", dataConn);
                    comm.ExecuteNonQuery();
                }*/

                string origId = NoQuotes(toks[cols["ORIGIN_AIRPORT_ID"]]);
                /*if (!airports.Contains(origId))
                {
                    airports.Add(origId);
                    comm = new NpgsqlCommand("INSERT INTO airport (airport_id,code)
Values(" + origId + ",'" + orig + "');", dataConn);
                    try
                    {
                        comm.ExecuteNonQuery();
                    }
                    catch (Exception ex) { }
                }*/

                string destId = NoQuotes(toks[cols["DEST_AIRPORT_ID"]]);
                /*if (!airports.Contains(destId))
                {
                    airports.Add(destId);
                    comm = new NpgsqlCommand("INSERT INTO airport (airport_id,code)
Values(" + destId + ",'" + dest + "');", dataConn);
                    try
                    {
                        comm.ExecuteNonQuery();
                    }
                    catch (Exception ex) { }
                }*/
                //string fltNum = NoQuotes(toks[cols["FL_NUM"]]);
                //string fltID = cc + fltNum.PadLeft(5, '0') + "_" + dateString + "_" + orig + "_" +
dest;
                int nSeats = 0;
                if(toks[cols["TAIL_NUM"]] != "\"\"")
                {
```

```
                    string seatQuery = "SELECT seats, acq_date FROM Aircraft WHERE
tailnumber LIKE '" + NoQuotes(toks[cols["TAIL_NUM"]]) + "%' AND year <= " +
toks[cols["YEAR"]] + " ORDER BY acq_date;";
                    DataTable dt = new DataTable();
                    NpgsqlDataAdapter da = new NpgsqlDataAdapter(seatQuery, dataConn);
                    da.Fill(dt);
                    DataRow[] rows = dt.Select("seats > 0","acq_date desc");
                    if (rows.Length > 0)
                       nSeats = Convert.ToInt16(rows[0]["seats"].ToString());
                 }
 string q1 = "INSERT INTO sw_flight
(flightid,airline,number,tailnumber,seats,date,origin,destination,scheddep,schedarr,distan
ce,schedelap,cancelled,diverted";
string v1 = " VALUES('" + fltID + "','" + cc + "'," + fltNum + ",'" +
NoQuotes(toks[cols["TAIL_NUM"]]) + "'," + nSeats.ToString() + ",'" +
NoQuotes(toks[cols["FL_DATE"]]) + "','" + orig + "','" + dest + "'," +
TimeString(toks[cols["CRS_DEP_TIME"]]) + ",";
v1 += TimeString(toks[cols["CRS_ARR_TIME"]]) + "," + toks[cols["DISTANCE"]] + "," +
toks[cols["CRS_ELAPSED_TIME"]] + ",B'" + toks[cols["CANCELLED"]].Substring(0, 1) + "',B'"
+ toks[cols["DIVERTED"]].Substring(0, 1) + "'";

if (toks[cols["CANCELLED"]].Substring(0, 1) != "1" && toks[cols["DIVERTED"]].Substring(0,
1) != "1")
                 {
                    string arrDelayNew;
                    if (!cols.ContainsKey("ARR_DELAY_NEW"))
                    {
                       int n = (int)Convert.ToDouble(toks[cols["ARR_DELAY"]]);
                       n = Math.Max(0, n);
                       arrDelayNew = n.ToString();
                    }
                    else
                    {
                       arrDelayNew = toks[cols["ARR_DELAY_NEW"]];
                    }
                    q1 +=
",actdep,actarr,depdiff,depdelay,arrdiff,arrdelay,taxiout,taxiin,wheelsoff,wheelson,actelap,
airtime";
                    v1 += "," + TimeString(toks[cols["DEP_TIME"]]) + "," +
TimeString(toks[cols["ARR_TIME"]]) + "," + toks[cols["DEP_DELAY"]] + "," +
toks[cols["DEP_DELAY_NEW"]] + "," + toks[cols["ARR_DELAY"]];

                    v1 += "," + arrDelayNew + "," + toks[cols["TAXI_IN"]] + "," +
```

```
toks[cols["TAXI_OUT"]] + "," + TimeString(toks[cols["WHEELS_OFF"]]) + "," +
TimeString(toks[cols["WHEELS_ON"]]);
                        v1 += "," + toks[cols["ACTUAL_ELAPSED_TIME"]] + "," +
toks[cols["AIR_TIME"]];
                    }
                    else
                    {
                        if (toks[cols["DIVERTED"]].Substring(0, 1) != "1")
                        {
                            q1 += ",cancelcode";
                            v1 += ",'" + toks[cols["CANCELLATION_CODE"]] + "'";
                        }
                    }
                    if (toks[cols["CARRIER_DELAY"]] != "")
                    {
q1 += ",carrierdelay,nasdelay,securitydelay,weatherdelay,lateaircraftdelay";
v1 += "," + toks[cols["CARRIER_DELAY"]] + "," + toks[cols["NAS_DELAY"]] + "," +
toks[cols["SECURITY_DELAY"]] + "," + toks[cols["WEATHER_DELAY"]] + "," +
toks[cols["LATE_AIRCRAFT_DELAY"]];
                    }
                    q1 += ")";
                    v1 += ");";
                    v1 = NoQuotes(v1);
                    q1 += v1;

                    try
                    {
                        comm = new NpgsqlCommand(q1, dataConn);
                        flights += comm.ExecuteNonQuery();
                    }
                    catch (Exception ex) { }
                }

                this.Invoke(tbcb, textBox3, airlines.Count.ToString());
                this.Invoke(tbcb, textBox6, airports.Count.ToString());
                this.Invoke(tbcb, textBox2, lines.ToString());
                this.Invoke(tbcb, textBox7, fltID);
                this.Invoke(tbcb, textBox4, flights.ToString());
                this.Invoke(pgcb);
            }
            sr.Close();
        }
        private static string TimeString(string t)
```

```
{
    string result = NoQuotes(t).PadLeft(4,'0');
    result = "\"" + result.Substring(0,2) + ":" + result.Substring(2) + "\"";
    return result;
}
private void UpdateTextBox(TextBox tb, string s)
{
    tb.Text = s;
    tb.Refresh();
}
private void UpdateProgress()
{
progressBar1.Value = Math.Min(progressBar1.Maximum,
Math.Max(progressBar1.Minimum, (int)((progressBar1.Maximum -
progressBar1.Minimum) * (bytesRead / (double)bytesToRead) +
progressBar1.Minimum)));
}
private static string NoQuotes(string s)
{
    return s.Replace("\"", "");
}
private static string[] GetCSVFields(string s)
{
    List<string> fields = new List<string>();
    string s2 = s.Substring(0);
    int sIndex = 0;
    int eIndex = 0;
    bool hasQuote = false;
    for (int i = 0; i < s2.Length; i++)
    {
        char c = s2[i];
        if (c == ',')
        {
            if (!hasQuote)
            {
                eIndex = i - 1;
                fields.Add(s2.Substring(sIndex, eIndex - sIndex + 1));
                sIndex = i + 1;
            }
        }
        if (c == '"')
        {
            hasQuote = !hasQuote;
```

```
            }
        }
        fields.Add(s2.Substring(sIndex, s2.Length - sIndex));
        return fields.ToArray();
    }
    private static string Elapsed(string s1, string s2)
    {
        s1 = NoQuotes(s1);
        s2 = NoQuotes(s2);
        string h1,m1,h2,m2;
        if (s1 == "" || s2 == "")
            return "0";
        if(s1.Contains(':'))
        {
            h1 = s1.Substring(0, s1.IndexOf(':'));
            m1 = s1.Substring(s1.IndexOf(':') + 1);
            h2 = s2.Substring(0, s1.IndexOf(':'));
            m2 = s2.Substring(s1.IndexOf(':') + 1);
        }
        else
        {
            h1 = s1.Substring(0,s1.Length-2);
            h2 = s2.Substring(0, s2.Length - 2);
            m1 = s1.Substring(s1.Length - 2);
            m2 = s1.Substring(s2.Length - 2);
        }
        return "0";
        int hr1 = Convert.ToInt16(h1);
        int min1 = Convert.ToInt16(m1);
        int hr2 = Convert.ToInt16(h2);
        int min2 = Convert.ToInt16(m2);

        TimeSpan el1 = new TimeSpan(hr1, min1, 0);
        TimeSpan el2 = new TimeSpan(hr2, min2, 0);
        TimeSpan el = el2.Subtract(el1);
        if(el.TotalMinutes < 0)
            el.Add(new TimeSpan(24,0,0));
return String.Concat(el.TotalHours.ToString("00"), ":", el.TotalMinutes.ToString("00"));
    }
  }
}
```

Flight Interval Calculation

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using Npgsql;
namespace FlightIntervalUpdater
{
  public partial class Form1 : Form
  {
    System.Threading.Thread mainThread;
    delegate void stringCallback(string s1, string s2, int n, int n2);
    delegate void progressCallback(int n, int n2);
    public Form1()
    {
      InitializeComponent();
    }

    private void button1_Click(object sender, EventArgs e)
    {
      mainThread = new System.Threading.Thread(MainLoop);
      mainThread.Start();
    }
    private void MainLoop()
    {
      NpgsqlConnection conn = new
NpgsqlConnection("Server=localhost;Port=5432;DataBase=postgres;User
Id=postgres;Password=9a55w0rd");
      conn.Open();

      DataTable dt = new DataTable();
      NpgsqlDataAdapter da = new NpgsqlDataAdapter("SELECT DISTINCT airline, airport
FROM monthly_summary2 ORDER BY airline", conn);
      da.Fill(dt);
      stringCallback updateLocation = new stringCallback(UpdateLocation);
      progressCallback updateCount = new progressCallback(UpdateProgress);
      for(int i =0; i<dt.Rows.Count; i++)
      {
        DataRow dr = dt.Rows[i];
```

```
            string arpt = dr["airport"].ToString();
            string arln = dr["airline"].ToString();
            try
            {
               this.Invoke(updateLocation, new object[] { arpt, arln, i, dt.Rows.Count });
            }
            catch (Exception ex) { }
            NpgsqlCommand comm1 = new NpgsqlCommand("SELECT flightid, date,
scheddep FROM flight WHERE airline ='" + arln + "' AND origin ='" + arpt + "' ORDER BY
date,scheddep ASC", conn);
            comm1.CommandTimeout = 3600;
            da = new NpgsqlDataAdapter(comm1);
            DataTable dt2 = new DataTable();
            try
            {
               da.Fill(dt2);
            }
            catch (Npgsql.NpgsqlException ex2)
            {
               while (MessageBox.Show(ex2.Message, "Command exception",
MessageBoxButtons.RetryCancel).Equals(DialogResult.Retry))
               {
                  da.Fill(dt2);
               }
               MessageBox.Show(String.Concat(arpt, ",", arln));
               return;
            }

            for (int j = 1; j < dt2.Rows.Count; j++)
            {
               DataRow dr2 = dt2.Rows[j];
               DataRow pdr = dt2.Rows[j-1];
               DateTime date = (DateTime) dr2["date"];
               DateTime time = (DateTime) dr2["scheddep"];
               date = date.Add(time.TimeOfDay);
               DateTime date2 = (DateTime) pdr["date"];
               DateTime time2 = (DateTime) pdr["scheddep"];
               date2 = date2.Add(time2.TimeOfDay);
               int minutes = (int)(date.Subtract(date2).TotalMinutes);
               string sql = "UPDATE flight SET lastdep = " + minutes.ToString("0") + " WHERE
airline ='" + arln + "' AND origin='" + arpt+ "' AND flightid = '" + dr2["flightid"].ToString() +
"'";
               NpgsqlCommand comm = new NpgsqlCommand(sql, conn);
```

```
                comm.ExecuteNonQuery();
                try
                {
                    this.Invoke(updateCount,new object[]{j+1,dt2.Rows.Count});
                }
                catch(Exception ex3){}
            }
        }
    }
    private void UpdateLocation(string s1, string s2, int n, int total)
    {
        textBox1.Text = s1;
        textBox2.Text = s2;

        progressBar1.Value =progressBar1.Minimum + (n*(progressBar1.Maximum -
progressBar1.Minimum)) / total;
        progressBar1.Refresh();

        textBox1.Refresh();
        textBox2.Refresh();
    }
    private void UpdateProgress(int n, int total)
    {
        textBox3.Text = string.Concat(n, "/", total);
        progressBar2.Value = progressBar2.Minimum + (n * (progressBar2.Maximum -
progressBar2.Minimum))/total;
        textBox3.Refresh();
        progressBar2.Refresh();
    }
  }
}
```

Appendix B: Select SQL Statements for Data Processing

Join Procedure

Step 1:

```
UPDATE flight
SET seats = aircraft.seats
FROM aircraft
WHERE aircraft.tailnumber = flight.tailnumber
        AND aircraft.year = flight.year
```

Step 2:

```
UPDATE flight
SET seats = aircraft.seats
FROM aircraft
WHERE flight.seats IS NULL
        AND aircraft.tailnumber = flight.tailnumber
        AND aircraft.year < flight.year
```

Summary

```
INSERT INTO mnthly_sumry (airline, month,mth,airport,departures,departures_w_seats,
departure_seats_per_flight,avg_departure_delay,peaking)
SELECT airline, month, month-((month/100)*100), origin, COUNT(*),COUNT(seats),
AVG(seats),AVG(avoidable_delay),STDDEV_SAMP(lastdep)/AVG(lastdep)
FROM flight
WHERE cancelled <> B'1'
GROUP BY month,airline,origin;

(define destination_summary)
SELECT flight.airline, flight.month, flight.destination, count(*) AS cnt,
   count(flight.seats) AS sts, avg(flight.seats) AS spf,
   avg(flight.netarrdelay) AS dly
  FROM flight
 WHERE flight.cancelled <> B'1'::"bit"
 GROUP BY flight.month, flight.airline, flight.destination;

UPDATE mnthly_sumry
SET arrivals = destination_summary.cnt,
        arrivals_w_seats = destination_summary.sts,
        arrival_seats_per_flight = destination_summary.spf ,
        avg_arrival_delay = destination_summary.dly
FROM destination_summary
WHERE mnthly_sumry.airline = destination_summary.airline
        AND mnthly_sumry.month = destination_summary.month
```

```
	AND mnthly_sumry.airport = destination_summary.destination
UPDATE mnthly_sumry
SET est_passengers = (arrival_seats_per_flight*arrivals + departure_seats_per_flight*
departures)/2,
	total_delay = avg_arrival_delay + avg_departure_delay,
	total_seats = (arrival_seats_per_flight *arrivals_w_seats +
departure_seats_per_flight*departures_w_seats)/(arrivals_w_seats+departures_w_seats
),
total_flights = arrivals + departures
```

Change Calculation

```
UPDATE mnthly_sumry as m1
SET flight_change = m2.total_flights - m1.total_flights
FROM mnthly_sumry as m2
WHERE m2.airline = m1.airline
	AND m2.airport = m1.airport
	AND m2.month = m1.month +100
```

Group Delineation

```
UPDATE fit_results SET flight_group = 'study' WHERE flightP < .05 AND flightM <> 0
```

Dominance

```
UPDATE fit_results
SET dominance = fit_results.total_flights/t1.all_flights
FROM
(SELECT airport, SUM(total_flights) AS all_flights
FROM fit_results
GROUP BY airport)AS t1
WHERE fit_results.airport = t1.airport
```

Appendix C: Select R scripts

Regresssions

```
>for(i in 1:length(siteList$airline)) get_fit(siteList[i,1],siteList[i,2])

#script to look up, do fit and store fit information for airport airline pair
get_fit <- function(airline, airport )
{
	status <- paste("Processed", airline, airport, sep=" ")
	query<-paste("select
total_delay,seat_change,peak_change,flight_change,avl_change from mnthly_sumry
where airline ='", airline, "' AND airport ='", airport, "' AND (seat_change IS NOT NULL
AND flight_change IS NOT NULL)",sep="")
	info <- dbGetQuery(con, query)

	if(length(info$total_delay) < 2) return("Not enough rows")

	fit1 <- lm(info$seat_change~info$total_delay)
	seatM <- fit1$coefficients[2]
	seatP <- lm_p_value(fit1)
	if(!is.finite(seatM)) seatM<-0.0
	if(!is.finite(seatP)) seatP<-1.0

	fit2 <- lm(info$flight_change~info$total_delay)
	flightM <- fit2$coefficients[2]
	flightP <- lm_p_value(fit2)
	if(!is.finite(flightM)) flightM<-0.0
	if(!is.finite(flightP)) flightP<-1.0

	fit3 <- lm(info$peak_change~info$total_delay)
	peakM <- fit3$coefficients[2]
	peakP <- lm_p_value(fit3)
	if(!is.finite(peakM)) peakM<-0.0
	if(!is.finite(peakP)) peakP<-1.0

	fit4 <- lm(info$avl_change~info$total_delay)
	avlM <- fit4$coefficients[2]
	avlP <- lm_p_value(fit4)
	if(!is.finite(avlM)) avlM<-0.0
	if(!is.finite(avlP)) avlP<-1.0

	insert1 <- paste("insert into fit_results
(airline,airport,count,seatm,seatp,flightm,flightp,peakm,peakp,avlm,avlp,avg_delay)
```

```
VALUES('", airline, "','", airport, "',",sep="")
          insert2 <-
paste(length(info$flight_change),seatM,seatP,flightM,flightP,peakM,peakP,avlm,avlp,mea
n(info$total_delay,na.rm=T),sep=",")
          insert3 <- paste(insert1,insert2,")",sep="")
          dbSendQuery(con,insert3)

          return(status)
}
#function to pull out pertinent variable from lm model object
lm_p_value <- function (modelobject) {
   if (class(modelobject) != "lm") stop("Not an object of class 'lm' ")
   f <- summary(modelobject)$fstatistic
   p <- pf(f[1],f[2],f[3],lower.tail=F)
   attributes(p) <- NULL
   return(p)
}
```

Delay T-Tests

```
groups <- dbGetQuery(con,"SELECT avg_delay, avl_group, seat_group, flight_group,
peak_group FROM fit_results WHERE count >= 60")
> t.test(groups$avg_delay~groups$seat_group)

      Welch Two Sample t-test

data:  groups$avg_delay by groups$seat_group
t = 0.4539, df = 68.8, p-value = 0.6513
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8880132  1.4110761
sample estimates:
mean in group control   mean in group study
          15.37811            15.11658

> t.test(groups$avg_delay~groups$flight_group)

      Welch Two Sample t-test

data:  groups$avg_delay by groups$flight_group
t = -0.2429, df = 108.52, p-value = 0.8086
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.313035  1.026379
sample estimates:
```

```
mean in group control   mean in group study
         15.34258            15.48591

> t.test(groups$avg_delay~groups$peak_group)

      Welch Two Sample t-test

data:  groups$avg_delay by groups$peak_group
t = -2.1905, df = 76.59, p-value = 0.03153
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.0871780 -0.1469273
sample estimates:
mean in group control   mean in group study
         15.21726            16.83431

> t.test(groups$avg_delay~groups$avl_group)

      Welch Two Sample t-test

data:  groups$avg_delay by groups$avl_group
t = -0.6174, df = 108.432, p-value = 0.5382
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.5365483  0.8066313
sample estimates:
mean in group control   mean in group study
         15.31703            15.68199
```

Delay T-Tests by Airline

```
for(i in 1:length(airlineList$airline)) arlnTTest(airlineList[i,1])

#t-tests by airline
arlnTTest <-function(airline)
{
         equalVar<-F
         query <- paste("SELECT avg_delay, avl_group, seat_group, flight_group,
peak_group FROM fit_results WHERE airline ='",airline,"' AND count >= 60",sep="")
         groups <- dbGetQuery(con,query)
         if(length(groups$avg_delay)<2) return("failed")

         if(length(unique(groups$seat_group)) != 2 |
length(groups$seat_group[groups$seat_group=="study"]) < 2)
```

```
	{
		seatp <- 1
		seatcm <-0
		seatsm <-0
	}else
	{
		seatTest<-
t.test(groups$avg_delay~groups$seat_group,var.equal=equalVar,alternative="less")
		seatp <- seatTest$p.value
		seatcm <-seatTest$estimate[1]
		seatsm <- seatTest$estimate[2]
	}

	if(length(unique(groups$flight_group)) != 2|
length(groups$flight_group[groups$flight_group=="study"]) < 2)
	{
		flightp <- 1
		flightcm <- 0
		flightsm <-0
	}else
	{
		flightTest<-
t.test(groups$avg_delay~groups$flight_group,var.equal=equalVar,alternative="less")
		flightp <- flightTest$p.value
		flightcm<-flightTest$estimate[1]
		flightsm <- seatTest$estimate[2]

	}

	if(length(unique(groups$peak_group)) != 2 |
length(groups$peak_group[groups$peak_group=="study"]) < 2)
	{
		peakp <- 1
		peakcm <- 0
		peaksm <- 0
	}else
	{
		peakTest<-
t.test(groups$avg_delay~groups$peak_group,var.equal=equalVar,alternative="less")
		peakp <- peakTest$p.value
		peakcm <-peakTest$estimate[1]
		peaksm <- peakTest$estimate[2]
```

```
	}

	if(length(unique(groups$avl_group)) != 2 |
length(groups$avl_group[groups$avl_group=="study"]) < 2)
	{
		avlp <- 1
		avlcm <- 0
		avlsm <- 0
	}else
	{
		avlTest<-
t.test(groups$avg_delay~groups$avl_group,var.equal=equalVar,alternative="less")
		avlp <- avlTest$p.value
		avlcm <-avlTest$estimate[1]
		avlsm <- avlTest$estimate[2]

	}

	insert<-paste("INSERT INTO airline_summary2
(airline,seatp,seatcm,seatsm,flightp,flightcm,flightsm,peakp,peakcm,peaksm,avlp,avlcm,a
vlsm) VALUES('",airline,"'",sep="")
	insert<-
paste(insert,seatp,seatcm,seatsm,flightp,flightcm,flightsm,peakp,peakcm,peaksm,avlp,avl
cm,avlsm,sep=",")
	insert<-paste(insert,")",sep="")
	dbSendQuery(con,insert)
	return(airline)
}
```

Dominance T-Tests

```
groups <- dbGetQuery(con,"SELECT dominance, avl_group, seat_group, flight_group,
peak_group FROM fit_results WHERE count >= 60")
> t.test(groups$dominance~groups$seat_group)

	Welch Two Sample t-test

data:  groups$dominance by groups$seat_group
t = 2.6886, df = 76.549, p-value = 0.008801
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.01823215 0.12239045
sample estimates:
mean in group control   mean in group study
```

```
          0.2061290           0.1358177

> t.test(groups$dominance~groups$flight_group)

        Welch Two Sample t-test

data:  groups$dominance by groups$flight_group
t = 2.3143, df = 128.251, p-value = 0.02224
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.008351247 0.106810840
sample estimates:
mean in group control   mean in group study
           0.207657            0.150076

> t.test(groups$dominance~groups$peak_group)

        Welch Two Sample t-test

data:  groups$dominance by groups$peak_group
t = 0.5533, df = 82.935, p-value = 0.5816
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.04656588  0.08245631
sample estimates:
mean in group control   mean in group study
          0.2025928           0.1846476

> t.test(groups$dominance~groups$avl_group)

        Welch Two Sample t-test

data:  groups$dominance by groups$avl_group
t = 2.4565, df = 130.669, p-value = 0.01534
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.01163339 0.10788767
sample estimates:
mean in group control   mean in group study
          0.2079083           0.1481478
```

Rerun seats test without monopoly destinations

```
> groups <- dbGetQuery(con,"SELECT dominance, avl_group, seat_group, flight_group,
```

```
peak_group FROM fit_results WHERE count >= 60 AND dominance < 1")
> t.test(groups$dominance~groups$seat_group)

	Welch Two Sample t-test

data:  groups$dominance by groups$seat_group
t = 0.8456, df = 66.423, p-value = 0.4008
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02903409  0.07170416
sample estimates:
mean in group control   mean in group study
           0.1571527             0.1358177
```

Appendix D: List of Airport Abbreviations

Abbreviation	Name	Location
ABE	Lehigh Valley International	Allentown/Bethlehem/Easton, PA
ABI	Abilene Regional	Abilene, TX
ABQ	Albuquerque International Sunport	Albuquerque, NM
ABR	Aberdeen Regional	Aberdeen, SD
ABY	Southwest Georgia Regional	Albany, GA
ACK	Nantucket Memorial	Nantucket, MA
ACT	Waco Regional	Waco, TX
ACV	Arcata	Arcata/Eureka, CA
ACY	Atlantic City International	Atlantic City, NJ
ADK	Adak NS	Adak Island, AK
ADQ	Kodiak Airport	Kodiak, AK
AEX	Alexandria International	Alexandria, LA
AGS	Augusta Regional at Bush Field	Augusta, GA
AKN	King Salmon Airport	King Salmon, AK
ALB	Albany International	Albany, NY
ALO	Waterloo Regional	Waterloo, IA
AMA	Amarillo International	Amarillo, TX
AMA	Rick Husband Amarillo International	Amarillo, TX
ANC	Ted Stevens Anchorage International	Anchorage, AK
APF	Naples Municipal	Naples, FL
ART	Watertown International	Watertown, NY
ASE	Aspen Pitkin County Sardy Field	Aspen, CO
ATL	Atlanta Municipal	Atlanta, GA
ATL	William B. Hartsfield Atlanta International	Atlanta, GA
ATL	Hartsfield-Jackson Atlanta International	Atlanta, GA
ATW	Outagamie County Regional	Appleton, WI
AUS	Austin - Bergstrom International	Austin, TX
AVL	Asheville Regional	Asheville, NC
AVP	Wilkes Barre Scranton International	Scranton/Wilkes-Barre, PA

AZO	Kalamazoo/Battle Creek International	Kalamazoo, MI
BDL	Bradley International	Hartford, CT
BET	Bethel Airport	Bethel, AK
BFL	Meadows Field	Bakersfield, CA
BGM	Greater Binghamton/Edwin A. Link Field	Binghamton, NY
BGR	Bangor International	Bangor, ME
BHM	Birmingham-Shuttlesworth International	Birmingham, AL
BIL	Billings Logan International	Billings, MT
BIS	Bismarck Municipal	Bismarck/Mandan, ND
BJI	Bemidji/Beltrami County	Bemidji, MN
BKG	Branson Airport	Branson, MO
BLI	Bellingham International	Bellingham, WA
BMI	Central Illinois Regional	Bloomington/Normal, IL
BNA	Nashville International	Nashville, TN
BOI	Boise Air Terminal	Boise, ID
BOS	Logan International	Boston, MA
BPT	Jack Brooks Regional	Beaumont/Port Arthur, TX
BQK	Brunswick Golden Isles	Brunswick, GA
BQN	Rafael Hernandez	Aguadilla, PR
BRO	Brownsville South Padre Island International	Brownsville, TX
BRW	Wiley Post/Will Rogers Memorial	Barrow, AK
BTM	Bert Mooney	Butte, MT
BTR	Baton Rouge Metropolitan/Ryan Field	Baton Rouge, LA
BTV	Burlington International	Burlington, VT
BUF	Buffalo Niagara International	Buffalo, NY
BUR	Hollywood-Burbank Midpoint	Burbank, CA
BWI	Baltimore/Washington International	Baltimore, MD
BZN	Bozeman Yellowstone International	Bozeman, MT
CAE	Columbia Metropolitan	Columbia, SC
CAK	Akron-Canton Regional	Akron, OH
CDC	Cedar City Regional	Cedar City, UT

CDV	Merle K Mudhole Smith	Cordova, AK
CEC	Jack McNamara Field	Crescent City, CA
CHA	Lovell Field	Chattanooga, TN
CHO	Charlottesville Albemarle	Charlottesville, VA
CHS	Charleston AFB/International	Charleston, SC
CIC	Chico Municipal	Chico, CA
CID	Cedar Rapids Municipal	Cedar Rapids/Iowa City, IA
CKB	North Central West Virginia	Clarksburg/Fairmont, WV
CLD	McClellan-Palomar	Carlsbad, CA
CLE	Cleveland-Hopkins International	Cleveland, OH
CLL	Easterwood Field	College Station/Bryan, TX
CLT	Charlotte Douglas International	Charlotte, NC
CMH	Port Columbus International	Columbus, OH
CMI	University of Illinois/Willard	Champaign/Urbana, IL
CMX	Houghton County Memorial	Hancock/Houghton, MI
COD	Yellowstone Regional	Cody, WY
COS	Peterson Field	Colorado Springs, CO
COU	Columbia Regional	Columbia, MO
CPR	Casper/Natrona County International	Casper, WY
CRP	Corpus Christi International	Corpus Christi, TX
CRW	Yeager	Charleston/Dunbar, WV
CSG	Columbus Metropolitan	Columbus, GA
CVG	Cincinnati/Northern Kentucky International	Cincinnati, OH
CWA	Central Wisconsin	Mosinee, WI
CYS	Cheyenne Regional/Jerry Olson Field	Cheyenne, WY
DAB	Daytona Beach International	Daytona Beach, FL
DAL	Dallas Love Field	Dallas, TX
DAY	James M Cox/Dayton International	Dayton, OH
DBQ	Dubuque Regional	Dubuque, IA
DCA	Ronald Reagan Washington National	Washington, DC
DEN	Stapleton International	Denver, CO

DFW	Dallas/Fort Worth International	Dallas/Fort Worth, TX
DHN	Dothan Regional	Dothan, AL
DLG	Dillingham Airport	Dillingham, AK
DLH	Duluth International	Duluth, MN
DRO	Durango La Plata County	Durango, CO
DSM	Des Moines Municipal	Des Moines, IA
DTW	Detroit Metro Wayne County	Detroit, MI
EAU	Chippewa Valley Regional	Eau Claire, WI
ECP	Northwest Florida Beaches International	Panama City, FL
EGE	Eagle County Regional	Eagle, CO
EKO	Elko Regional	Elko, NV
ELM	Elmira/Corning Regional	Elmira/Corning, NY
ELP	El Paso International	El Paso, TX
ERI	Erie International/Tom Ridge Field	Erie, PA
EUG	Mahlon Sweet Field	Eugene, OR
EVV	Evansville Regional	Evansville, IN
EWN	Craven County Regional	New Bern/Morehead/Beaufort, NC
EWR	Newark Liberty International	Newark, NJ
EYW	Key West International	Key West, FL
FAI	Fairbanks International	Fairbanks, AK
FAR	Hector International	Fargo, ND
FAT	Fresno Yosemite International	Fresno, CA
FAY	Fayetteville Regional/Grannis Field	Fayetteville, NC
FCA	Glacier Park International	Kalispell, MT
FLG	Flagstaff Pulliam	Flagstaff, AZ
FLL	Fort Lauderdale-Hollywood International	Fort Lauderdale, FL
FLO	Florence Regional	Florence, SC
FNT	Bishop International	Flint, MI
FSD	Joe Foss Field	Sioux Falls, SD
FSM	Fort Smith Regional	Fort Smith, AR
FWA	Fort Wayne International	Fort Wayne, IN

GCC	Gillette Campbell County	Gillette, WY
GEG	Spokane International	Spokane, WA
GFK	Grand Forks International	Grand Forks, ND
GGG	East Texas Regional	Longview, TX
GGG	East Texas Regional	Longview, TX
GJT	Grand Junction Regional	Grand Junction, CO
GLH	Greenville Municipal	Greenville, MS
GNV	Gainesville Regional	Gainesville, FL
GPT	Gulfport-Biloxi International	Gulfport/Biloxi, MS
GRB	Austin Straubel International	Green Bay, WI
GRI	Grand Island Air Park	Grand Island, NE
GRK	Robert Gray AAF	Killeen, TX
GRR	Kent County	Grand Rapids, MI
GRR	Gerald R. Ford International	Grand Rapids, MI
GSO	Piedmont Triad International	Greensboro/High Point, NC
GSP	Greenville-Spartanburg International	Greer, SC
GST	Gustavus Airport	Gustavus, AK
GTF	Great Falls International	Great Falls, MT
GTR	Golden Triangle Regional	Columbus, MS
GUC	Gunnison-Crested Butte Regional	Gunnison, CO
HDN	Yampa Valley	Hayden, CO
HHH	Hilton Head Airport	Hilton Head, SC
HKY	Hickory Regional	Hickory, NC
HLN	Helena Regional	Helena, MT
HNL	Honolulu International	Honolulu, HI
HOB	Lea County Hobbs	Hobbs, NM
HOU	William P Hobby	Houston, TX
HPN	Westchester County	White Plains, NY
HRL	Valley International	Harlingen/San Benito, TX
HSV	Huntsville International-Carl T Jones Field	Huntsville, AL
HTS	Tri-State/Milton J. Ferguson Field	Ashland, WV

HVN	Tweed New Haven	New Haven, CT
IAD	Washington Dulles International	Washington, DC
IAH	George Bush Intercontinental/Houston	Houston, TX
ICT	Wichita Mid-Continent	Wichita, KS
IDA	Idaho Falls Regional	Idaho Falls, ID
ILG	Greater Wilmington	Wilmington, DE
ILM	Wilmington International	Wilmington, NC
IND	Indianapolis International	Indianapolis, IN
INL	Falls International	International Falls, MN
IPL	Imperial County	El Centro, CA
ISO	Kinston Regional Jetport at Stallings Field	Kinston, NC
ISP	Long Island MacArthur	Islip, NY
ITH	Ithaca Tompkins Regional	Ithaca/Cortland, NY
ITO	Hilo International	Hilo, HI
IYK	Inyokern-Kern County	Inyokern, CA
JAC	Jackson Hole	Jackson, WY
JAN	Jackson - Evers International	Jackson/Vicksburg, MS
JAX	Jacksonville International	Jacksonville, FL
JFK	John F. Kennedy International	New York, NY
JNU	Juneau International	Juneau, AK
KOA	Kona International Airport at Keahole	Kona, HI
KTN	Ketchikan International	Ketchikan, AK
LAN	Capital Region International	Lansing, MI
LAS	McCarran International	Las Vegas, NV
LAW	Lawton-Fort Sill Regional	Lawton/Fort Sill, OK
LAX	Los Angeles International	Los Angeles, CA
LBB	Lubbock Preston Smith International	Lubbock, TX
LCH	Lake Charles Regional	Lake Charles, LA
LEX	Blue Grass	Lexington, KY
LFT	Lafayette Regional	Lafayette, LA
LGA	LaGuardia	New York, NY

LGB	Long Beach Airport	Long Beach, CA
LIH	Lihue Airport	Lihue, HI
LIT	Bill and Hillary Clinton Nat Adams Field	Little Rock, AR
LMT	Kingsley Field	Klamath Falls, OR
LNK	Lincoln Airport	Lincoln, NE
LRD	Laredo AFB	Laredo, TX
LRD	Laredo International	Laredo, TX
LSE	La Crosse Municipal	La Crosse, WI
LWB	Greenbrier Valley	Lewisburg, WV
LWS	Lewiston Nez Perce County	Lewiston, ID
LYH	Lynchburg Regional/Preston Glenn Field	Lynchburg, VA
MAF	Midland International	Midland/Odessa, TX
MBS	Tri City	Saginaw/Bay City/Midland, MI
MBS	MBS International	Saginaw/Bay City/Midland, MI
MCI	Kansas City International	Kansas City, MO
MCN	Middle Georgia Regional	Macon, GA
MCO	Orlando International	Orlando, FL
MDT	Harrisburg International	Harrisburg, PA
MDW	Chicago Midway International	Chicago, IL
MEI	Key Field	Meridian, MS
MEM	Memphis International	Memphis, TN
MFE	McAllen Miller International	Mission/McAllen/Edinburg, TX
MFR	Rogue Valley International - Medford	Medford, OR
MGM	Montgomery Regional	Montgomery, AL
MHK	Manhattan Regional	Manhattan/Ft. Riley, KS
MHT	Manchester Airport	Manchester, NH
MIA	Miami International	Miami, FL
MKC	Kansas City Downtown	Kansas City, MO
MKC	Charles B. Wheeler Downtown	Kansas City, MO
MKE	General Mitchell International	Milwaukee, WI
MKG	Muskegon County	Muskegon, MI

MLB	Melbourne International	Melbourne, FL
MLI	Quad City International	Moline, IL
MLU	Monroe Regional	Monroe, LA
MMH	Mammoth Lakes Airport	Mammoth Lakes, CA
MOB	Mobile Regional	Mobile, AL
MOD	Modesto City-County-Harry Sham Field	Modesto, CA
MOT	Minot International	Minot, ND
MQT	Sawyer International	Marquette, MI
MRY	Monterey Peninsula	Monterey, CA
MSN	Dane County Regional-Truax Field	Madison, WI
MSO	Missoula International	Missoula, MT
MSP	Minneapolis-St Paul International	Minneapolis, MN
MSY	Louis Armstrong New Orleans International	New Orleans, LA
MTH	The Florida Keys Marathon	Marathon, FL
MTJ	Montrose Regional	Montrose/Delta, CO
MVY	Martha's Vineyard Airport	Martha's Vineyard, MA
MWH	Grant County International	Moses Lake, WA
MYR	Myrtle Beach International	Myrtle Beach, SC
OAJ	Albert J Ellis	Jacksonville/Camp Lejeune, NC
OAK	Metropolitan Oakland International	Oakland, CA
OGG	Kahului Airport	Kahului, HI
OKC	Will Rogers World	Oklahoma City, OK
OMA	Eppley Airfield	Omaha, NE
OME	Nome Airport	Nome, AK
ONT	Ontario International	Ontario, CA
ORD	Chicago O'Hare International	Chicago, IL
ORF	Norfolk International	Norfolk, VA
OTH	North Bend Municipal	North Bend/Coos Bay, OR
OTZ	Ralph Wien Memorial	Kotzebue, AK
OXR	Oxnard	Oxnard/Ventura, CA
PAH	Barkley Regional	Paducah, KY

PBI	Palm Beach International	West Palm Beach/Palm Beach, FL
PDX	Portland International	Portland, OR
PFN	Bay County	Panama City, FL
PHF	Patrick Henry International	Newport News/Williamsburg, VA
PHL	Philadelphia International	Philadelphia, PA
PHX	Phoenix Sky Harbor International	Phoenix, AZ
PIA	General Downing - Peoria International	Peoria, IL
PIA	General Downing - Peoria International	Peoria, IL
PIE	St. Petersburg-Clearwater International	St. Petersburg, FL
PIH	Pocatello Regional	Pocatello, ID
PIR	Pierre Municipal	Pierre, SD
PIT	Pittsburgh International	Pittsburgh, PA
PLN	Pellston Regional Airport of Emmet County	Pellston, MI
PMD	Palmdale USAF Plant 42	Palmdale, CA
PNS	Pensacola Regional	Pensacola, FL
PSC	Tri Cities	Pasco/Kennewick/Richland, WA
PSE	Mercedita	Ponce, PR
PSG	Petersburg James A Johnson	Petersburg, AK
PSP	Palm Springs International	Palm Springs, CA
PUB	Pueblo Memorial	Pueblo, CO
PVD	Theodore Francis Green State	Providence, RI
PWM	Portland International Jetport	Portland, ME
RAP	Rapid City Regional	Rapid City, SD
RDD	Redding Municipal	Redding, CA
RDM	Roberts Field	Bend/Redmond, OR
RDU	Raleigh-Durham International	Raleigh/Durham, NC
RFD	Chicago/Rockford International	Rockford, IL
RHI	Rhinelander/Oneida County	Rhinelander, WI
RIC	Richmond International	Richmond, VA
RKS	Rock Springs Sweetwater County	Rock Springs, WY
RNO	Reno/Tahoe International	Reno, NV

ROA	Roanoke Regional/Woodrum Field	Roanoke, VA
ROC	Greater Rochester International	Rochester, NY
ROW	Roswell International Air Center	Roswell, NM
RST	Rochester Municipal	Rochester, MN
RSW	Southwest Florida International	Fort Myers, FL
SAF	Santa Fe Municipal	Santa Fe, NM
SAN	San Diego International Lindbergh Fl	San Diego, CA
SAT	San Antonio International	San Antonio, TX
SAT	San Antonio International	San Antonio, TX
SAV	Savannah/Hilton Head International	Savannah, GA
SBA	Santa Barbara Municipal	Santa Barbara, CA
SBN	South Bend Airport	South Bend, IN
SBP	San Luis Obispo County Regional	San Luis Obispo, CA
SCC	Deadhorse Airport	Deadhorse, AK
SCE	State College Air Depot	State College, PA
SDF	Louisville International-Standiford Field	Louisville, KY
SEA	Seattle/Tacoma International	Seattle, WA
SFO	San Francisco International	San Francisco, CA
SGF	Springfield-Branson National	Springfield, MO
SGU	St George Municipal	St. George, UT
SHV	Shreveport Regional	Shreveport, LA
SIT	Sitka Rocky Gutierrez	Sitka, AK
SJC	San Jose International	San Jose, CA
SJT	San Angelo Regional/Mathis Field	San Angelo, TX
SJU	Luis Munoz Marin International	San Juan, PR
SLC	Salt Lake City International	Salt Lake City, UT
SLE	McNary Field	Salem, OR
SMF	Sacramento International	Sacramento, CA
SMX	Santa Maria Public/Capt. G. Allan Hancock Field	Santa Maria, CA
SNA	John Wayne Airport-Orange County	Santa Ana, CA
SOP	Moore County	Pinehurst/Southern Pines, NC

SPI	Capital	Springfield, IL
SPI	Abraham Lincoln Capital	Springfield, IL
SPN	Francisco C. Ada Saipan International	Saipan, TT
SPS	Sheppard AFB/Wichita Falls Municipal	Wichita Falls, TX
SRQ	Sarasota/Bradenton International	Sarasota/Bradenton, FL
STL	Lambert-St. Louis International	St. Louis, MO
STT	Cyril E King	Charlotte Amalie, VI
STX	Alexander Hamilton	Christiansted, VI
STX	Henry E. Rohlsen	Christiansted, VI
SUN	Friedman Memorial	Sun Valley/Hailey/Ketchum, ID
SUX	Sioux Gateway/Col. Bud Day Field	Sioux City, IA
SWF	Stewart International	Newburgh/Poughkeepsie, NY
SYR	Syracuse Hancock International	Syracuse, NY
TEX	Telluride Regional	Telluride, CO
TLH	Tallahassee Regional	Tallahassee, FL
TOL	Toledo Express	Toledo, OH
TPA	Tampa International	Tampa, FL
TRI	Tri-Cities Regional TN/VA	Bristol/Johnson City/Kingsport, TN
TTN	Trenton Mercer	Trenton, NJ
TUL	Tulsa International	Tulsa, OK
TUP	Tupelo Regional	Tupelo, MS
TUS	Tucson International	Tucson, AZ
TVC	Cherry Capital	Traverse City, MI
TWF	Joslin Field - Magic Valley Regional	Twin Falls, ID
TXK	Texarkana Regional-Webb Field	Texarkana, AR
TYR	Tyler Pounds Regional	Tyler, TX
TYS	McGhee Tyson	Knoxville, TN
UTM	Tunica Municipal	Tunica, MS
VLD	Valdosta Regional	Valdosta, GA
VPS	Northwest Florida Regional	Valparaiso, FL
WRG	Wrangell Airport	Wrangell, AK

WYS	Yellowstone	West Yellowstone, MT
XNA	Northwest Arkansas Regional	Fayetteville, AR
XNA	Northwest Arkansas Regional	Fayetteville, AR
YAK	Yakutat Airport	Yakutat, AK
YKM	Yakima Air Terminal/McAllister Field	Yakima, WA
YUM	Yuma MCAS/Yuma International	Yuma, AZ

Appendix E: List of Carrier Abbreviations

Code	Carrier Name
9E	Pinnacle Airlines Inc.
AA	American Airlines Inc.
AS	Alaska Airlines Inc.
B6	JetBlue Airways
CO	Continental Air Lines Inc.
DL	Delta Air Lines Inc.
EV	ExpressJet Airlines Inc.
F9	Frontier Airlines Inc.
FL	AirTran Airways Corporation
HA	Hawaiian Airlines Inc.
MQ	American Eagle Airlines Inc.
NW	Northwest Airlines Inc.
OH	Comair Inc.
OO	SkyWest Airlines Inc.
TZ	American Trans Air Inc.
UA	United Air Lines Inc.
US	US Airways Inc.
WN	Southwest Airlines Co.
XE	ExpressJet Airlines Inc.
YV	Mesa Airlines Inc.

Appendix F: Study Groups

Format

Airline identifier: (airport identifier) (airport identifier) . . .

Seats per Flight

AA : MCI SAN SNA

AS : ANC DCA LGB ONT PHX SEA TUS

B6 : HOU SRQ

CO : HNL PDX PHX RNO SEA SMF

DL : DAB JAC MDT ORD TLH

EV : ATL BDL BTV CHS EWN FWA GSP ISP MDW MGM PSP PWM ROC

F9 : ANC PHL

FL : ATL BTV FNT MIA

KH : OAK

MQ : CHS FNT GPT GRB LFT LSE

NW : DFW FAR FNT LAN LIT MBS MCI MDT MHT MSP PVD TPA TVC

OH : BGR BNA CRW ILM STL SYR

OO : AVP DFW HPN IAH RAP TUS

UA : ABQ BWI DEN DFW DTW JFK LAX MCI PDX PHL PIT SAN SEA SFO SMF

US : ABQ CMH DTW EWR MHT SEA SLC

WN : BUR

XE : ALB AVL BDL CLE CLT

YV : BDL BHM DFW ELP EUG EWR LGB ONT PHX PNS SAT

Flights

AA : FLL JFK LAX MCO PDX SAN SJU TPA

AS : LAS OAK PDX SAN SJC SMF SNA

B6 : ORD

CO : DEN FLL IND MCI ORD PIT

DL : JAC LAS LAX MKE OAK PHX

EV : EWN GTR IND MCN MEM MSP OAJ ORF SBN SGF

F9 : CAK DAY LGA SEA

FL : HPN MDW TPA

HA : LIH SJC

MQ : BDL CID CLL DBQ GSO IAD LFT LIT SAV SFO

NW : ALB DFW HDN JAX MDW MHT MSP MSY ROC RSW SBN TPA

OH : ATL AVL BUF CHA DEN EWR GRB IAD ORD PVD ROC SGF TRI

OO : BHM CPR EKO GEG IAH LAS MBS ORF PIH PMD RNO TVC TWF

UA : AUS DFW DTW FSD ICT LAX LGA OAK PDX SAT SEA SFO SJC SMF SNA

US : CMH DEN DFW DTW FLL JFK MIA OAK PBI PHX PIT RDU RNO ROC RSW SRQ TUS

WN : MAF

XE : ALB AMA BDL BTV CLE DAB HSV LEX MEM ORD ORF PIT PWM RIC SHV SYR XNA

YV : BNA CLD DRO ELP EWR GSO HNL ICT ITO KOA PHX RIC SBA

Peaking

AA : CMH EWR FLL LGA RDU SNA

AS : EWR FAI HNL LAS LGB OME SEA SJC SMF

B6 : LAX

CO : BOS CMH IAH IND MCI OGG ORD

DL : ANC GRB HDN JFK LAS LAX MKE MSN ONT ORD

EV : GTR HHH MSP OAJ ROA

F9 : LGA MSY RSW

FL : HPN MCI PHX

HA : PHX

MQ : AMA BDL BNA CID CLL CMH GSO LAS LSE ORF SAV SFO

NW : ANC AZO CLT DFW DTW HDN LGA MDW MSP MSY ORD RSW SJC TPA

OH : ATL CAK CID IAD PBI ROC

OO : BHM CDC CHS CPR EKO FNT FWA GEG IAH MKG SBN SPI TWF

UA : AUS BUR DAY DFW ICT LAX LGA OAK PVD RDU SJC TUS

US : BNA CLT CMH EGE ELP EWR JFK MSY PIT RNO ROC RSW SRQ

WN : BDL LIT MAF

XE : ALB AMA BTV CID CLE ELP HSV LEX MEM ORD ORF PIT PVD RIC SFO SYR XNA

YV : CLD GRR GSO GUC ITO JFK MSN OKC SFO TRI

Available Seats

AA : BWI FLL JFK LAX MCO PDX SJU TPA

AS : LAS OAK PDX SAN SJC SMF SNA

B6 : ORD

CO : DEN FLL IND MCI ORD PIT

DL : LAS LAX MKE OAK PHX

EV : EWN GNV IND MCN MEM MSP ORF SBN SGF

F9 : CAK DAY LGA SEA

FL : HPN TPA

HA : LIH SJC

MQ : BDL CID DBQ GSO IAD LIT SAV SFO

NW : HDN HNL JAX MDW MSP MSY ROC RSW SAN SBN TPA

OH : ATL AVL BUF CHA DEN EWR GRB IAD ORD PVD ROC SGF TRI

OO : BHM CPR EKO EUG GEG IAH LAS LAX MBS MCI ORF PMD SAN TVC TWF

UA : AUS DFW DTW FSD ICT LAX LGA OAK PDX SEA SJC SMF SNA

US : CMH DEN DFW FLL HNL JFK MIA MSP OAK PBI PHL PHX PIT RDU RNO ROC RSW SJU SRQ TUS

WN : MAF

XE : ALB AMA BDL BTV CLE DAB HSV LEX MEM ORD ORF PIT PWM RIC SHV SYR XNA

YV : BNA CLD DRO ELP EWR GSO HNL ICT ITO KOA RIC

References

Bhadra, Dipasis, Jacqueline Kee(2008) Structure and dynamics of the core US air travel markets: A basic empirical analysis of domestic passenger demand. *Journal of Air Transport Management*, 14, 1, 27-39.

Britto, Rodrigo , Martin Dresner, Augusto Voltes(2012) The impact of flight delays on passenger demand and societal welfare. *Transportation Research Part E: Logistics and Transportation Review*, 48 2, 460-469.

Brueckner, J. K. (2003). Airline Traffic and Urban Economic Development. *Urban Studies*, 40(8), 1455-1469.

Brueckner, J. K. (2004), Network Structure and Airline Scheduling. The Journal of Industrial Economics, 52:291–312

Brueckner, Jan K. and Kurt Van Dender(2008)Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate, *Journal of Urban Economics*, 64, 2, 288-295.

Button, Kenneth, Somik Lall(1999)The economics of being an airport hub city, *Research in Transportation Economics*,5,75-105.

Costa, Tiago F.G., Gui Lohmann, Alessandro V.M. Oliveira(2010) A model to identify airport hubs and their importance to tourism in Brazil, *Research in Transportation Economics*,26,1, 3-11.

Daniel, Joseph I., Katherine Thomas Harback(2009) Pricing the major US hub airports, *Journal of Urban Economics*, 66, 1, 33-56.

Debbage, Keith G., Dawn Delk(2001) The geography of air passenger volume and local employment patterns by US metropolitan core area: 1973–1996, *Journal of Air Transport Management*,7,3,159-167.

Debbage, Keith G.(2002)AIRPORT RUNWAY SLOTS: Limits to Growth. *Annals of Tourism Research*, 29, 4, 933-951.

Derudder, Ben, Lomme Devriendt(2010) Frank Witlox, A spatial analysis of multiple airport cities, *Journal of Transport Geography*, 18, 3,345-353.

Derudder, Ben, Frank Witlox(2009)The impact of progressive liberalization on the spatiality of airline networks: a measurement framework based on the assessment of hierarchical differentiation, *Journal of Transport Geography*, 17, 4,276-284

Diana, Tony(2009)Do market-concentrated airports propagate more delays than less concentrated ones? A case study of selected U.S. airports, *Journal of Air Transport Management*, 15, 6, 280-286.

Dobruszkes, Frédéric(2006)An analysis of European low-cost airlines and their networks, *Journal of Transport Geography*, 14,4, 249-264.

Evans, Antony, Andreas Schäfer(2011) The impact of airport capacity constraints on future growth in the US air transportation system, *Journal of Air Transport Management*,17, 5, 288-295

Fageda, Xavier and Fernández-Villadangos, Laura. (2009). "Triggering competition in the Spanish airline market: The role of airport capacity and low-cost carriers*." Journal Of Air Transport Management* 15, 36-40.

Federal Aviation Administration(2012)"Los Angeles International ILS or LOC Rwy 6L" AL-237(FAA). http://aeronav.faa.gov/d-tpp/1204/00237IL6L.PDF

John Ferguson, Abdul Qadar Kara, Karla Hoffman, Lance Sherry(2012 Corrected Proof in Press available online January 11) "Estimating domestic US airline cost of delay based on European model", *Transportation Research Part C: Emerging Technologies*

Flores-Fillol, Ricardo(2010) Congested hubs, *Transportation Research Part B: Methodological*, 44, 3, 358-370.

Forbes, Silke J.(2008) "The effect of air traffic delays on airline prices." *International Journal of Industrial Organization*, 26, 5, 1218-1232.

Francis, Graham, Ian Humphreys, Stephen Ison, Michelle Aicken(2006) Where next for low cost airlines? A spatial and temporal comparative study. *Journal of Transport Geography*,14, 2,83-94.

Fuellhart, Kurt(2003)Inter-metropolitan airport substitution by consumers in an asymmetrical airfare environment: Harrisburg, Philadelphia and Baltimore. *Journal of Transport Geography*,11, 4, 285-296.

Fuellhart, Kurt(2007)Airport catchment and leakage in a multi-airport region: The case of Harrisburg International, *Journal of Transport Geography*, 15,4,231-244.

Gelhausen, Marc C.(2011)Modeling the effects of capacity constraints on air travelers' airport choice. *Journal of Air Transport Management*, 17, 2, 116-119.

Ginieis, Matías, María-Victoria Sánchez-Rebull, Fernando Campa-Planas(2012) "The academic journal literature on air transport: Analysis using systematic literature review methodology", *Journal of Air Transport Management*, 19, 31-35.

Graham,Brian. (1999). Claire Guyer, Environmental sustainability, airport capacity and European air transport liberalization: irreconcilable goals?, *Journal of Transport Geography*, 7, 3, 165-180.

Grubesic, Tony H., Timothy C. Matisziw, Matthew A. Zook(2009) Spatio-temporal fluctuations in the global airport hierarchies, *Journal of Transport Geography*, 17,4, 264-275.

Grubesic, Tony, Matthew Zook(2007)A ticket to ride: Evolving landscapes of air travel accessibility in the United States, *Journal of Transport Geography*, 15, 6, 417-430.

Halpern, Nigel, Svein Bråthen(2011)Impact of airports on regional accessibility and social development. *Journal of Transport Geography*, 19, 6, 1145-1154.

Hansen Mark(2002) "Micro-level analysis of airport delay externalities using deterministic queuing models: a case study", *Journal of Air Transport Management*, 8, 2, 73-87.

Huber, Hans(2009) Comparing spatial concentration and assessing relative market structure in air traffic, *Journal of Air Transport Management*, 15,4, 184-194.

Johnson,Tracy and Ian Savage(2006)Departure delays, the pricing of congestion, and expansion proposals at Chicago O'Hare Airport, *Journal of Air Transport Management*, 12, 4, 182-190.

Liu, Zhi-Jun, Keith Debbage, Brendan Blackburn(2006)Locational determinants of major US air passenger markets by metropolitan area, *Journal of Air Transport Management*,12,6,331-341.

Loo, B. P., Ho, H. W., and Wong, S. C. (2005). An application of the continuous equilibrium modeling approach in understanding the geography of air passenger flows in a multi-airport region. *Applied Geography, 25*, 2, 169-199.

Nunn, S. (2005). Flight Plans for Development: Aviation Investments and Outputs in Nine Metropolitan Regions, 1990 to 2002. *Economic Development Quarterly, 19*, 4, 295-312.

Madas, Michael A., Konstantinos G. Zografos(2008)Airport capacity vs. demand: Mismatch or mismanagement?, *Transportation Research Part A: Policy and Practice*,42, 1, 203-226.

Martín, Juan Carlos, Concepción Román(2003) Hub location in the South-Atlantic airline market: A spatial competition game. *Transportation Research Part A: Policy and Practice*,37,10,865-888.

Matisziw, T.C., T.H. Grubesic, (2010) Evaluating locational accessibility to the US air transportation system, *Transportation Research Part A: Policy and Practice*, 44, 9,710-722.

Matisziw, T.C. T.H. Grubesic,(2011) A spatial analysis of air transport access and the essential air service program in the United States, *Journal of Transport Geography,* 19, 93-105.

McMillen, Daniel P. (2004)Airport expansions and property values: the case of Chicago O'Hare Airport, *Journal of Urban Economics*, 55, 3,627-640.

.

Morrison, Steven A. ,Clifford Winston(2008) The effect of FAA expenditures on air travel delays, *Journal of Urban Economics*, 63, 2,669-678.

NEXTOR et al.(2010) "Total Delay Impact Study A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States".

http://catsr.ite.gmu.edu/pubs/FAA_TDI_Report_Final_10_6_10.pdf.

Pai, Vivek(2010) "On the factors that affect airline flight frequency and aircraft size" *Journal of Air Transport Management*, 16, 4,169-177.

Pantazis, Nadine, Ingo Liefner(2006) The impact of low-cost carriers on catchment areas of established international airports: The case of Hanover Airport, Germany. *Journal of Transport Geography*, 14, 4,265-272.

Pyrgiotis,Nikolas, Kerry M. Malone, and Amedeo Odoni(2012) "Modelling delay propagation within an airport network." *Transportation Research Part C: Emerging Technologies.* In press. Available online 25 June 2011.

Redondi, Renato, Paolo Malighetti, Stefano Paleari(2011).Hub competition and travel times in the world-wide airport network, *Journal of Transport Geography*, 19,6,1260-1271.

Reynolds-Feighan, Aisling(2001)Traffic distribution in low-cost and full-service carrier networks in the US air transportation market. *Journal of Air Transport Management*,7,5,265-275.

Santos,Georgina and Maël Robin(2010) "Determinants of delays at European airports." *Transportation Research Part B: Methodological.* 44, 3, 392-403.

Schumer, C.E.(2008) "Flight Delays Cost Passengers, Airlines and the US Economy Billions", Joint Committee Majority Staff.

Swan,William M(2002) "Airline route developments: a review of history." *Journal of Air Transport Management*, 8, 5, 349-353.

Van Dender, Kurt(2007)Determinants of fares and operating revenues at US airports. *Journal of Urban Economics*, 62, 2, 317-336.

Wilken, Dieter, Peter Berster, Marc C. Gelhausen(2011)New empirical evidence on airport capacity utilisation: Relationships between hourly and annual air traffic volumes, *Research in Transportation Business & Management*, 1,1, 118-127.