Final Presentation Report: Flight Delays in Relation to Airport Hubs

Abstract

This report describes a set of interactive and descriptive visualizations to analyze delays and their causes in the United States domestic commercial aviation market. The visualization was created in Tableau and consist of a story composed by 25 slides. The story first covers the most important aspects of flight traffic such as network topologies and airport connectedness. It then analyzes delays and breaks them down by different combinations of factors, showing the main causes. Finally, the visualizations show that airport locations where major airline hubs are based are the ones where most of the delays are concentrated and that the main contributor to the problem is simply airport congestion as evidenced by taxiing patterns.

Introduction and Motivation

The United States has an extensive air transportation market. Due to the geography of the United States and the generally large distances between major cities, air transportation is the preferred method of travel for trips over 300 miles (480 km), such as for business travelers and long distance vacation travelers. With exception for the Northeast Rail Line, air transportation is the first choice for intercity travel. In 2015 the US domestic air transportation market saw more than 690,000 Passenger Enplanements, 8,980,000 passengers flown and 5,589,000 Revenue Aircraft Miles Flown. Most of the traffic was concentrated in the top 45 airports by passengers, with Atlanta (ATL), Chicago (ORD), Los Angeles (LAX), Dallas (DFW) and New York (JFK) occupying the first five positions in the ranking of busiest airports.

The domestic US aviation transportation system encounters frequent delays. In 2015, roughly 18% of all flights were delayed. The airlines and Bureau of Transportation Statistics keep records for each flight delay and assign one of five categories. In addition, airlines can be differentiated by their size, network model, and market segments. As a consequence, airports can also be classified in terms of different features. Delays might be related to high traffic volumes and network interconnectedness, but understanding the relation among the different factors in play is a complex undertaking requiring the use of multiple visualizations.

Dataset Description

The visualization references the "2015 Flight Delays and Cancellations" dataset found at https://www.kaggle.com/usdot/flight-delays/data. The core dataset comprises 3 tables: airlines, airports, and flights.

- Airlines key attributes: IATA Code, Airline
- Airports key attributes: Airport, City, State, Country, Latitude, Longitude, IATA Code

- Flights key attributes: Year, Month, Day, Airline, Flight Number, Origin Airport, Destination Airport, Schedule Departure Time, Departure Delay, Taxi Out, Scheduled Time, Elapsed Time, Scheduled Arrival, Arrival Time, Arrival Delay, Diverted, Cancelled, Air System Delay, Security Delay, Airline Delay, Late Aircraft Delay, Weather Delay.

Two Jupyter notebooks were used to transform and augment the data: new.ipynb and moredatawrangling.ipynb

new.ipynb:

The Airports file is augmented with new fields: Hub Carriers, Focus Carriers, Hub Count, Focus Count. The file is output back to csv with the same name, replacing the original. Values for Hub Carriers, Focus Carriers, Hub Count, Focus Count fields were researched on the internet and added manually to the new file.

The flights file is augmented with the following fields: Airline Name, Total Delay, Early Flag, Total Taxi and output to csv with file name "dataset_flights.csv"

moredatawrangling.ipynb:

In order to create map path lines in Tableau, the Flights file was denormalized so that each destination and origin are represented as separate rows; a new column identifying Path ID was created by concatenating Origin and Destination Iata Codes in a string value. Additionally, a Path Order column was created to define whether the row references a departure (given value of "1") or an arrival (value of "2"). The file is joined with the augmented airports dataset and then output to csv as "routes.csv"

Additional Notes

The flights file was initially filtered so that only data for January 1 - 5 was kept. In total, this produced a file with 79,025 rows (each representing a flight between two locations).

Problem Statement / Task Description

Every year, the top 45 US airports account for the wide majority of passengers and movements. The below diagram shows Domestic flight data for the top 5 airports by passengers.

Source: http://www.aci.aero/ (All Data in Thousands)

Location	IATA Code	Passengers	Movements
Atlanta	ATL	101,491	882
Chicago	ORD	76,949	875
Los Angeles	LAX	74,937,004	655
Dallas	DFW	64,074	681
New York	JFK	56,827	438

Airlines are generally divided between traditional and budget carriers. Delta, American Airlines, and United are prime examples of traditional airlines while Southwest, Frontier and Spirit are examples of the latter. From an operational perspective, one of the main differentiators is the transport topology optimization methodology followed. Traditional airlines use a spoke-hub distribution paradigm in which traffic routes are organized as a series of 'spokes' that connect outlying points to a central 'hub.' Budget carriers instead use a point to point transit system in which each point has a direct route to every other point.

The top airports illustrated in the table at page 1 are all hubs for at least one major US airline. For example, American Airlines served 18.5 million passengers in Dallas DFW alone in 2016. With such a complex and highly interconnected system, delays are inevitable. In 2015 roughly 18% of all US domestic flights were delayed while 1.5% were cancelled.

The Bureau of Transportation Statistics and airlines categorize delays as follows:

- **Air Carrier:** The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
- **Extreme Weather:** Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
- **National Aviation System (NAS):** Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
- **Late-arriving aircraft:** A previous flight with same aircraft arrived late, causing the present flight to depart late.
- **Security:** Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

The visualization as a whole is intended as an interactive exploratory tool aimed at audiences that have no to basic knowledge of the aviation industry. Rather than relying on complex metrics, domain specific calculations, and industry terms, the story leverages interactive maps and highly visual idioms to highlight key factors, correlations and causality, ultimately making the learning process easy and intuitive.

The Tableau story is to be seen as a sequence with initial idioms approaching the problem from a higher level and subsequently narrowing down on the root cause of delays. Ultimately, the tasks provide a solution to the problem of showing that hub airports are highly congested and they account for most of the delays across the country. Specifically, the main task can be divided into "story phases" - each accomplishing a set of subtasks:

1. Show air transportation network topology and demonstrate that most of the traffic is concentrated in certain key hubs.
    - Subtasks: show route network map, flight destinations map, airport accessibility and destination variety.
2. Show that delay causes are not correlated among one another. Show that there is no clear delay cause across all airports. Instead, at key hubs delays are highly correlated with taxiing times and they are not as highly correlated to other delay causes.

- Delay Exploration Subtasks: show map of delays, weighted sum of delays by airline, cancellations by airport, delay causes by airline, airports with highest delays, delays for hub airports, delays for hub airports by airline, correlation among delay causes.
- Delay Solution Subtasks: taxiing times, flight catch up, flight schedules at hub vs non hub airports, taxiing and delays correlation at hub vs non hub airports.

3. Show that key hubs account for significantly higher delays as a percentage of all delays, even after accounting for the fact that they make up for the majority of flights (i.e. weighting by traffic volumes).

In total, the above tasks are accomplished with 23 idioms. 5 are interactive maps while the rest are histograms, distribution histograms, stacked plots, regression lines, scatter plots, and heat map/tables.

Related Work

Research Paper 1:
Enhancing Visual Analysis of Network Traffic Using a Knowledge Representation
Ling Xiao* John Gerth Pat Hanrahan
Stanford University

The initial maps in the Tableau story plot airline routes onto a map. The routes clearly form a network with airports as nodes and routes (i.e. line connections between airports) as edges. The prime scope of including such network maps is to show that traditional airlines follow a hub and spoke model while budget airlines follow a point to point model. However, having multiple airlines each with hundreds of routes all plotted on the same map makes it hard to identify patterns. While the visualization does solve part of this problem by encoding airlines with colors and providing interactivity features such as filters to remove or add airlines and/or airports to the map, the viewer can only really begin to make sense of the network map if one or two airlines at a time are visualized. Thus, it is hard to visualize the whole network and infer ideas from the whole picture. In addition, one might decide to only view routes for a particular airline and learn that operations are focused on a particular hub airport, but after switching the airline in the filter a few times he or she might forget the basic insights gained with respect to a previously viewed airline network.

The research paper "Enhancing Visual Analysis of Network Traffic Using a Knowledge Representation" presents a cyclic network visualization system that facilitates visual analysis representation. The main steps of this system are the identification of a discovery pattern by the viewer, the definition of a model, and the updating of a knowledge base to visually store the derived insights.

The system is designed for the analysis of network flow data and is thus applicable to the flight route network map presented in the Tableau story introduction. The system uses a declarative event pattern language to define patterns using clauses composed of logical connectives and predicates. The authors demonstrate the system with an example for a web page loading pattern involving only two flows:

SimpleWebPageLoad(x,y) = identical_source_IP(x,y)
AND identical_destination_IP(x,y)
AND time_within_2_seconds(x,y)
AND (destination_port_80(x) AND destination_port_80(y))

The clause above defines a pattern as identified by the analysts by describing two flows from the same IP, to the same IP, within 2 seconds of each other, and both to port 80 (associated with HTTP traffic). A parallel as applied to the network route map could be:

FlightRoute(x,y) = origin_airport(x)
AND destination_airport(y)
AND after_3PM(x,y)
AND (airline_is_Delta(x,y) OR airline_is_American(x,y))

Where x and y could be any two airports in the network. The above query thus returns routes between airports x and y, where both departure and landing are after 3 PM, for Delta or American Airlines. With x = ATL and y = SAN the pattern would return the flight number for the only flight available from Atlanta to San Diego after 3 PM, which would be operated by Delta.
Another example shown below could be used to define an airport pattern based on carrier hubs:

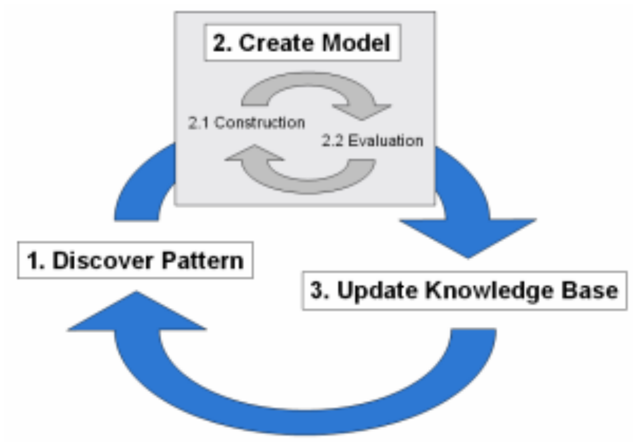
AirportHub(x,y,w,z) = airport_carrier(x)
AND hub_carrier_counts(y)
AND (state(w) OR state(z))

The above pattern would thus define all airports where Delta is a hub carrier and at least y hub carriers are present. If we use y=2 this pattern would return LAX and JFK.

After the pattern is defined, the authors' system envision the analyst building a model from the pattern discovered and then exploring the model space by iteratively constructing and evaluating candidate clauses; eventually converging to a clause that he or she believes captures the pattern. Visualizations may be used to identify patterns.

After selecting the pattern, the analyst is shown a list of predicates that describe the pattern along with the details of the selected events. To identify predicates that describe the pattern, each predicate in the knowledge base is tested on the selected events. Different strategies are used depending on predicate type. The system returns a unary predicate if it is satisfied by over 90% of the selected events. To create a clause describing the pattern, the analyst interactively constructs and evaluates candidate clauses.

Finally, once the model clauses are selected the analyst updates the knowledge base. The system will then create a higher level event representing each pattern instance in a new event table.



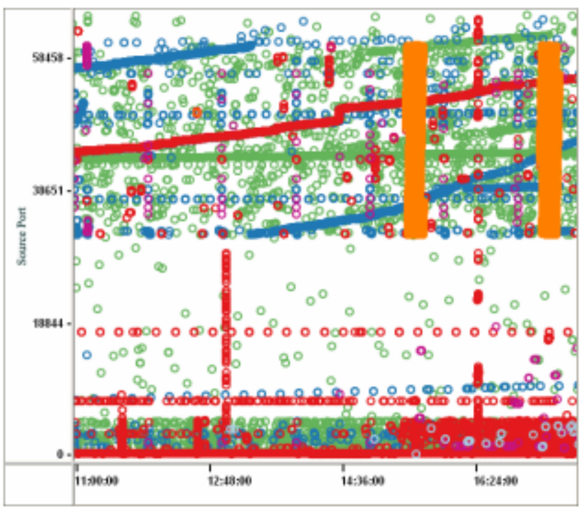


Figure 7. Event diagram showing mail (green), DNS (blue), scan (red), SSH logins (purple), SSH attacks (orange), and chat (indigo)

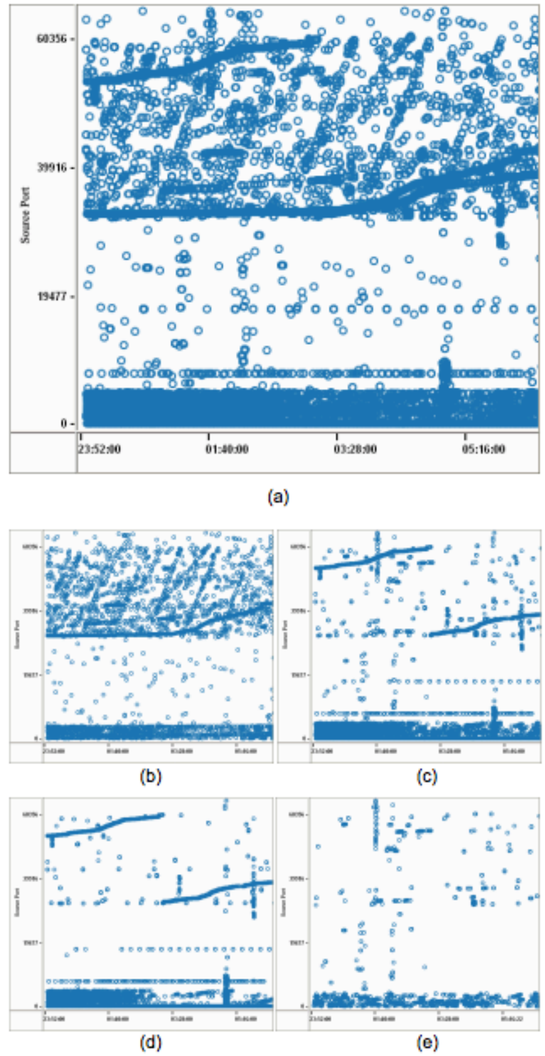


Figure 6. Event diagrams showing flows during residual analysis. (a) Original unidentified traffic (b) Flows with "mail" label (c) The residual after filtering out "mail" from Figure 6a. (d) Flows with the "scan" label (e) The residual after filtering out the "scan" label from Figure 6(c).

The example figures below show flow patterns as defined in the research paper example. The same type of flows could diagrams could be used to define route patterns for traditional airlines using a hub and spoke model. Additionally, it could be used to detect less evident patterns in budget airlines network routes, which despite not being officially based at certain hubs will still display route patterns.

Research Paper 2:
Metric-Based Network Exploration and Multiscale Scatterplot
Yves Chiricota - Universite' du Quebec a Chicoutimi, Canada
Fabien Jourdan and Guy Melancon - LIRMM UMR CNRS 5506, Montpellier, France

As part of the Tableau story many scatter plots take edge properties such as different delay types and plot them against one another in order to view how they behave with respect to each other (and detect whether any pattern is present). One problem encountered consists in the overcrowding of the plot as a result of the fact that 79,025 edges (flights) are in the network. As a means of proving that the edge properties vary based on the number of hub carriers present at the flight destination airport, the scatter plots are colored by this metric. Thus, flights directed to destinations such as LAX or ATL will be darker in color and can be isolated with a filter. However, this still leaves many points on the map (unless we restrict the range to airports with more than 2 hub carriers). This strongly limits the ability to learn from the scatter plots visualized. In fact, although points clustering around certain regions makes certain aspects of the network evident (e.g. flights to hub airports tend to have higher taxiing in times), the cluster simply becomes a "plateau" of points and nothing more than a high presence of points can be inferred. In addition, the color scheme intensity is hard to visualize as points colored with less intensity (representing flights directed to cities that have 0 hub carriers) will become less visible.

The research paper "Metric-Based Network Exploration and Multiscale Scatterplot" address this issue by presenting a methodology for visualizing large scale network data via scatter plots by means of imager processing techniques resulting in blurred clusters. The effects permit to encode the frequency of each cluster, and to select patches on the image as a way to extract sub-network in a more intuitive manner.

As an example, the authors plot a network linking the various Java classes of the "Resyn Assistant API" by assigning color to edges according to the degree of logical dependencies between classes (figure 2). Effectively the scatter plot is a tool to guide the exploration of the network.

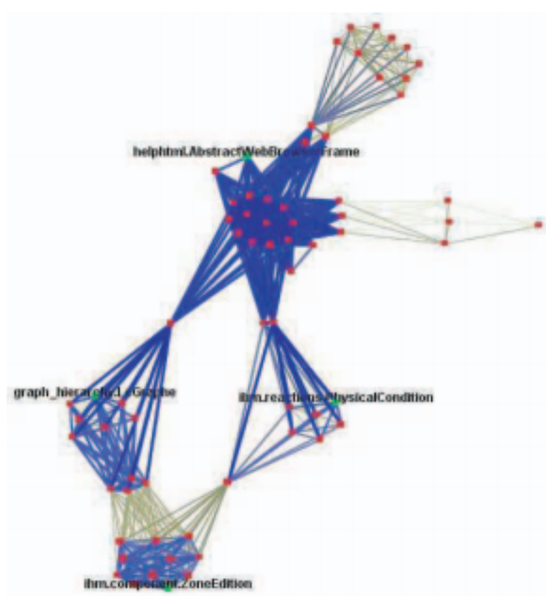


Figure 2: Example network : access graph induced from Java classes in the Resyn Assistant API. Edges have been coloured according to an application metric specific to Resyn.

Figure 3 below shows the same network on a scatter plot as it would have been plotted with traditional tools. Points of the scatterplot correspond to edges of the network where the x and y coordinates of the point are respectively given by the application specific metric and the strength metric associated with the edge. Points are also assigned varying (greyscale) intensity reflecting the frequencies since many edges are associated with the same (x, y) pair of values

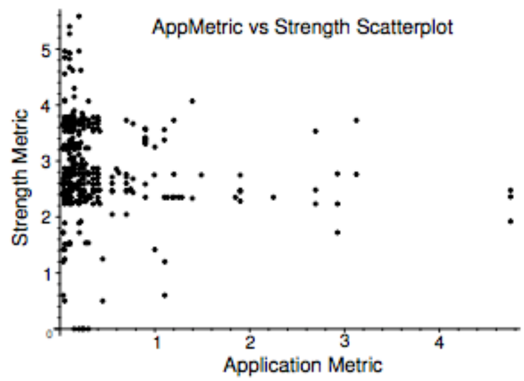


Figure 3: A 2D scatterplot displaying how the strength and application specific metric distribute over the edges of the network in Figure 2.

The majority of points in the scatterplot of Figure 3 are located on the left of the diagram and are organized into what seems to look like one or two bigger subgroups. But as the authors note, “This unsatisfactory description of the scatterplot actually points at an important aspect of metric-based exploration. **The user needs to be guided not only in her/his discovery of the network, but also when reading the plot in order to accomplish a more accurate and significant selection of subregions (and/or sub-networks)”.**

By means of Linear Algebra techniques, the image in Figure 3 is convoluted with a Gaussian kernel Gs produces a new image fs = Gs $*$ f , representing the original image after it has been blurred by a factor s (see figure 4 below). Figure 5 shows the same blurred scatterplot, but with a different parameter choice for s which results in the “borders” becoming wider and wider.

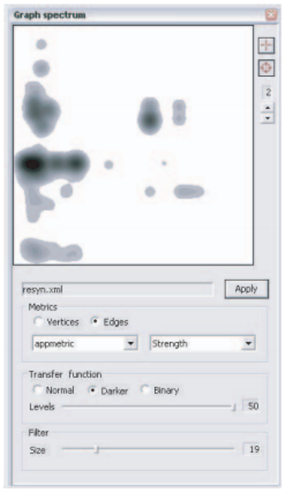


Figure 4: Convoluting the image in Figure 3 with a Gaussian filter produces a blurring effect and helps localize the regions of interest in the scatterplot.

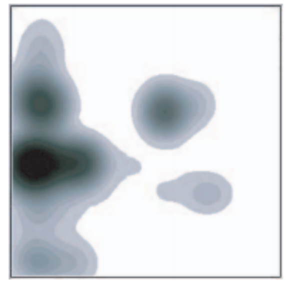


Figure 5: A greater value of s creates wider patches (when compared with Figure 4). The shape of the nested strips extends from the existing subregions.

The authors then showed that the application could be used to retrieve the nodes and edges as shown by the scatter plot to access the desired API's that are part of the Resyn application. First, they built a scatterplot computed over two metrics, one structural, the other contextual. The structural metric we used is the strength of edges indicating whether a link acts as a weak link between two distinct neighborhoods, or whether the edge sit at the core of a cluster. The Resyn specific metric is a weight measuring how much of the class (attributes and methods) is publicly accessible.

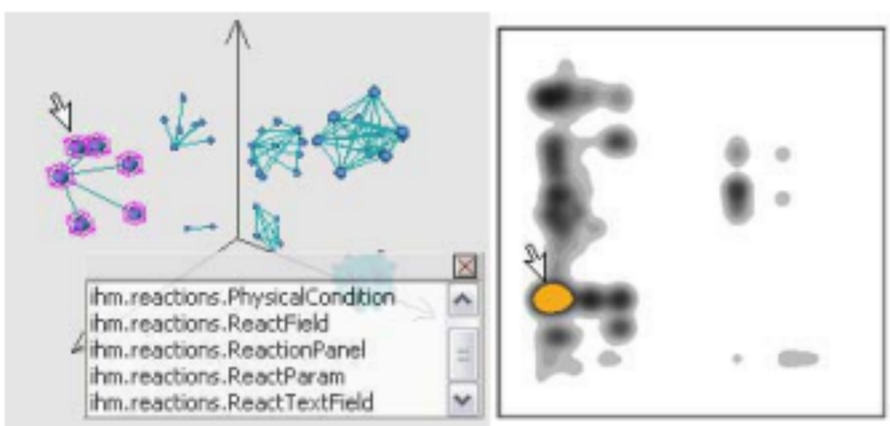


Figure 6: Darker patches in the scatterplot can be selected to recover the package structure of the Resyn Assistant API.

The selected region in figure 6 corresponds to edges simultaneously having an associated average strength value (y coordinate) and a low application specific metric value (x coordinate). Therefore, an edge with an average strength value necessarily sits in the middle of a moderately connected cluster. Contrary, the application metric

reflects how wide a variety a class services (public or protected attributes and methods) other classes and methods of the Rsyn API.
Thus, by selecting the yellow patch we identify classes that offer services to only a few other API members and that are part of moderately connected clusters. These are classes that are therefore less important in the application. These classes will live in the periphery of more densely connected clusters.

Solution - Visualization Design and Usage

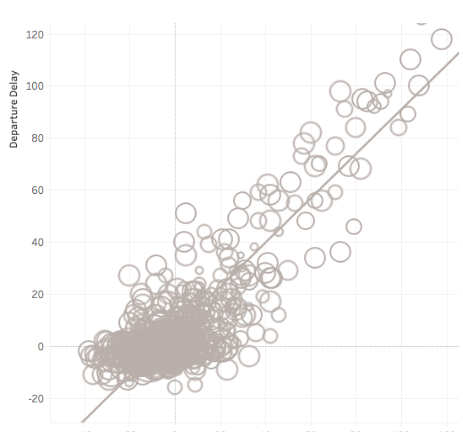
The design choices for the visualization as a whole have been driven by the need to display as much useful information in a sequential story starting from a general and high level problem approach to a more specific, narrow solution.

In order to place the problem into context and to grasp the magnitude of the dataset, it was necessary to start with multiple maps defining the route network for each airline. One of the fundamental factors into the problem definition is the existence of certain airport hubs where the majority fo flights are concentrated. While other approaches could have been used to visualize the flights dataset, maps were therefore the most informative.

Because the dataset includes more than 70,000 flights, there is a high risk of low separability and discriminability which has been addressed by means of interactive features mainly consisting of different filter combinations. Thus, one can reduce the amount of edges displayed in the graph to the bone by selecting individual origin and destination airports. While the network could have been extracted and displayed in more mathematical terms by using edge properties to define the space, maps are the most entertaining and intuitive choice for the intended audience.

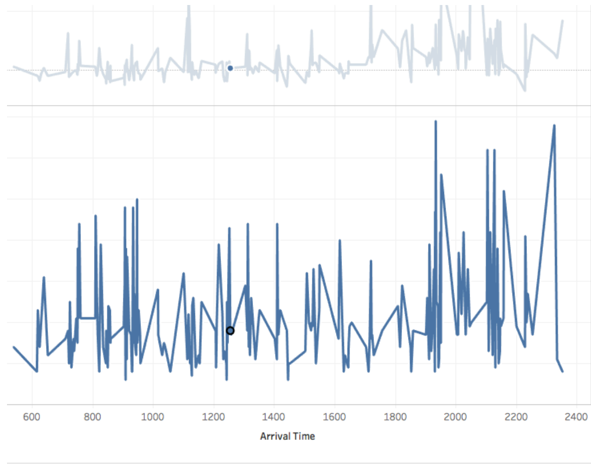
To address the second set of tasks related to understanding how delay factors relate among one another and to other variables, including multiple scatter plots along the same axis in the same idioms represents the best way to simultaneously search for patterns among potentially related variables. Once the initial scatter plots establish that hypothesized patterns do not exist, plot designs tend to include less variables at a time but focus on the ones that are inferred to be more important.

These later idioms are thus less complex but more narrowly address specific tasks. For example, one finding was that although many flights leave late, they arrive on time or early (this is because airlines always set a higher time window to each flight than really needed in order to give their pilots the opportunity to catch up in case of any departure delays). This is shown by idiom “Flight Catch Up” - shown below:



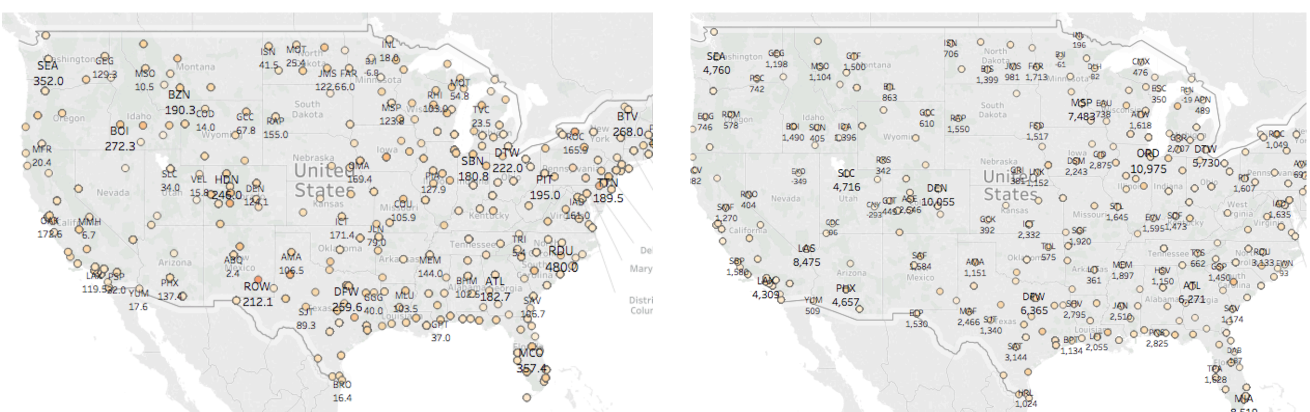
“Flight Catch Up” - Most flights are able to reduce any departure delays

As another example, an additional design choice is to "blow up" plots where certain patterns need to be highlighted - again this happens more frequently towards the end of the story.



"Non Hub Airport Arrival and Taxi In Delays"

Finally, to clearly make the point that hub airports disproportionately account for most delays, the very last idiom ("Conclusion - Map of Hub Airports with Weighted Sum Total") is conceived as a map with design choices that are very similar to the map in idiom "Average Total Delays Airport Map". This is done on purpose and in order to highlight that while the computation as made in a certain manner in the earlier stages of the story made things appear in a certain way, a reassessment as done after the findings explained in some of the last idioms proved that hub airports do indeed encounter more delays.



Left: "Average Total Delays Airport Map" - Right: "Conclusion - Map of Hub Airports with Weighted Sum Total" (Notice higher level of pop-out is used for hub airports)

In summary, application usage is focused around using separate interactive features as one progresses through the story to learn the key point.

Results

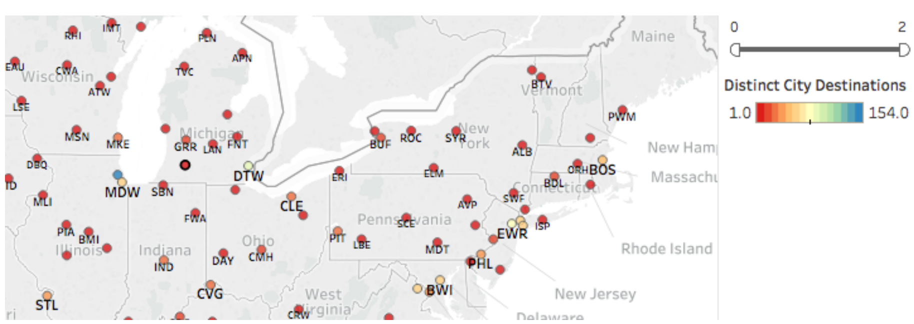
The visualization is composed by 23 idioms including maps, scatter plots, heat tables, line plots, and more. In terms of expressiveness, the visualization as a whole includes all the variables and information needed to accomplish the task at hand. The visualization as a whole is very expressive as it employs a wide variety of encodings to accomplish the tasks at hand. Ordered data is consistently displayed in an ordered fashion.

The visualization as whole is also very effective. In terms of channels chosen, all the idioms representing ordinal data use a position on a common scale methodology.

In terms of Accuracy, text or marker size along with color are used following a power law to emphasize data points that are more salient. As an example, idiom “Taxiing Times by Airports” shows all airports as circles on a 2D axis. A larger circle will automatically be closer to red and define airports with higher traffic volumes.

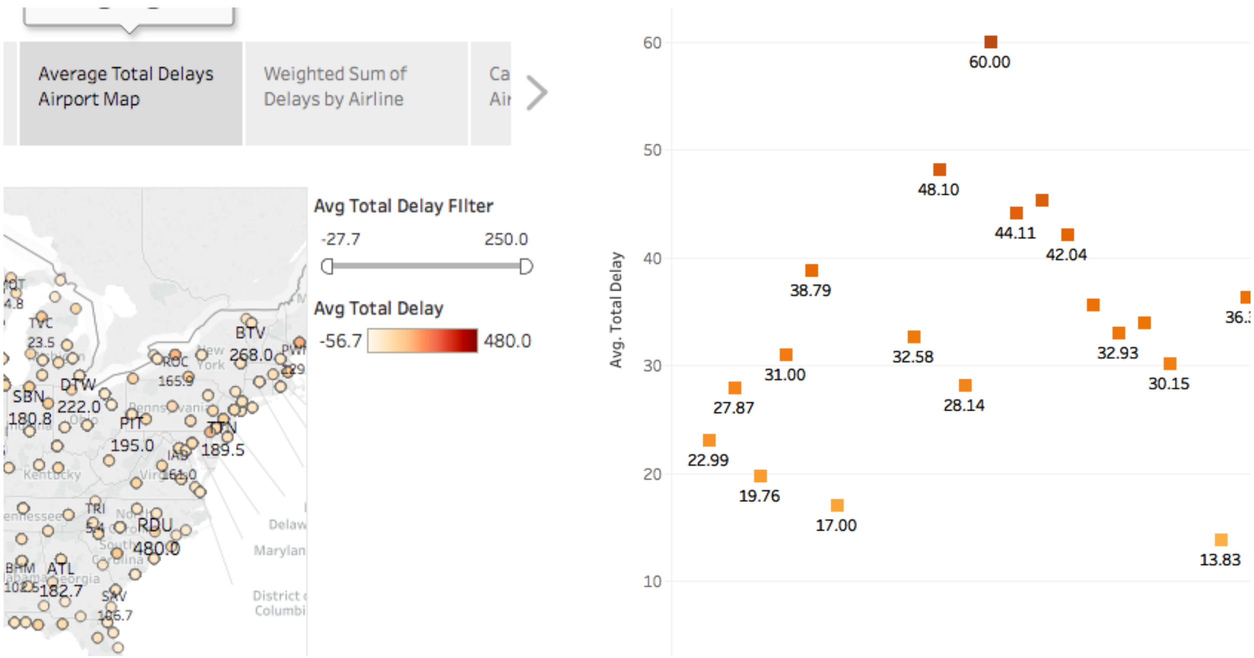
Accuracy through Marker Size and Color

As a second example, accuracy is defined by size in conjunction with color in the “Airport Accessibility” idiom to highlight airports that are more accessible to other cities and states. In this case, higher the levels of airport accessibility reflect in larger text label next to each airport circle on the map while also being displayed in a color closer to blue if more accessible, and red if more isolated.



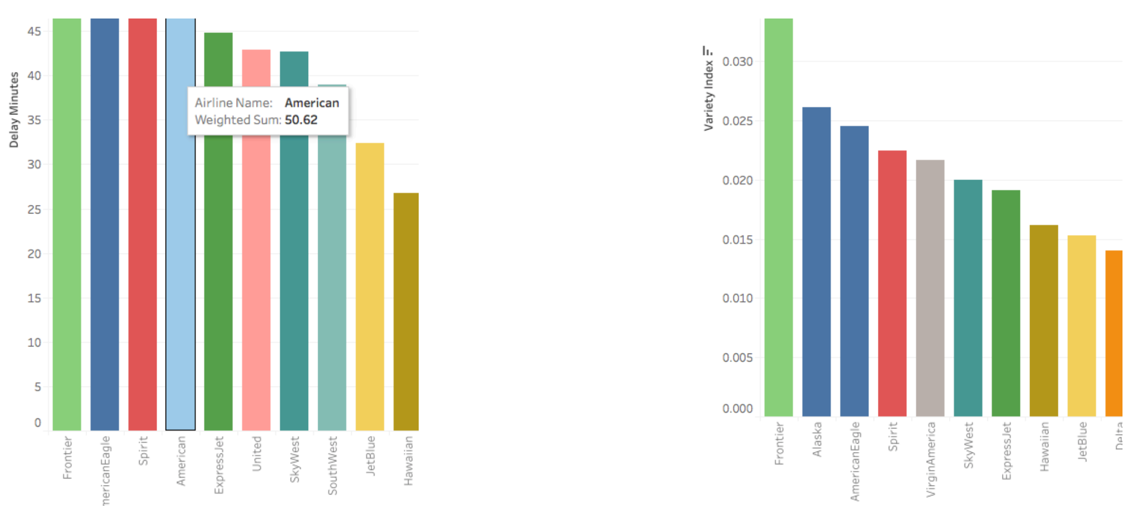
Accuracy defined by text label size and circle color in “Airport Accessibility”

Consistency is also maintained from a color encoding perspective. For example, when filtering airports for the number of hub carriers present, the same filter is used across multiple idioms. Color hue is used to defined degree of delay in idioms plotting delays such as maps. The same color encoding scheme is used across many idioms.



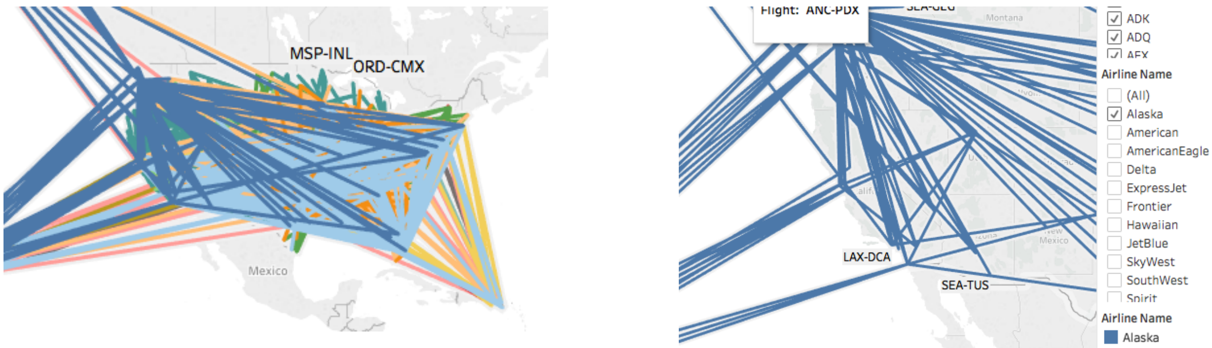
Comparison between Color Schemes in Idioms 5 and 16

As another example of color consistency across idioms, the same color coding is applied to airlines in idioms 17 and 19:



Comparison between Color Schemes in Idioms 17 and 19

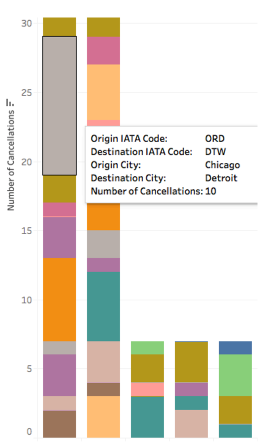
Discriminability across idioms including most maps is generally high as they are never too crowded. aAn exception to this might be the very first map in the story ("Route Network Map") which showcases the airline networks. Fortunately, the interactive filter comes to the rescue as it allows to remove airlines and their respective connections (network edges) from the map.



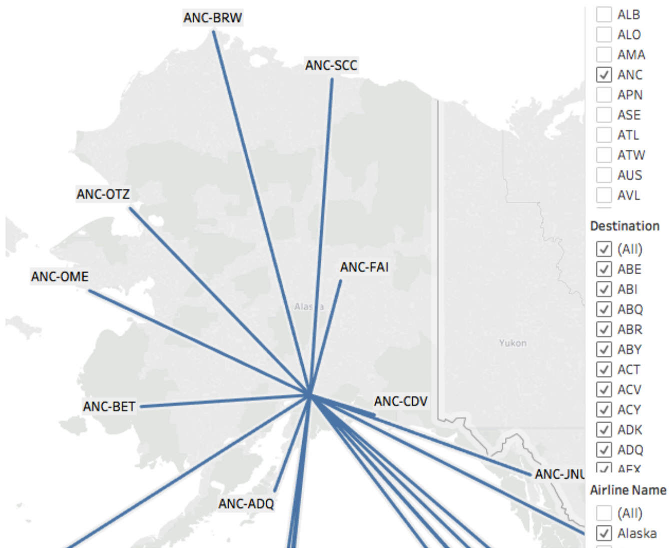
Route Network Map Idiom - Before and After Filter for Alaska Airlines

Separability is also high across all idioms but an exception is represented by the scatter plots displayed in the "Delay Cause Correlation Analysis" idiom where a circle for each of the 79,025 flights is shown with a small circle. Although a clear cluster takes shape near the origin (in most subplots), there may be more patterns (such as subgraphs) not directly visible as a result of the overcrowding of the plot. Luckily, these plots are only meant to be used to detect patterns between two variables at a time so the degree of discriminability is not a concern to accomplish the task at hand.

Degree of Pop-out is high whenever possible. As previously mentioned, the size of text labels and markers in certain idioms reflects channel salience and by consequence results in higher level of pop-out. As an additional example of this effectiveness factor, idiom "Cancellations by Airport" offers a visualization of all the cancellations occurred during the five day time frame for the ten airports (origins) with highest counts of cancellations. Each bar represents each of the origin airports and is divided into rectangles representing each destination airport hit by the cancellation. By hovering over each rectangle, the viewer can learn the name of the destination airport, origin airport, and the number of cancellations taken place. The higher this metric is, the larger the rectangle. The y - axis on the left can then be used to obtain a sum of all cancellations by origin airport and compare with others (notice all bars are sorted by total count so the origin airports with more cancellations will appear closest to the axis). As we can see in the screenshot to the left, Chicago ORD ranked highest in terms of cancellations as origin airport; among cancelled flights scheduled to depart from there, Detroit DTW was the destination with the highest count of cancelled flights. In other words, 10 ORD - DTW fights were cancelled.

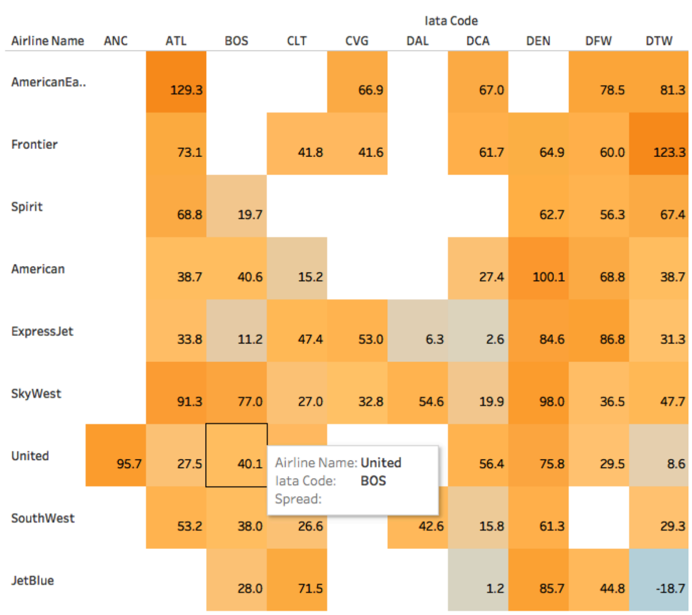


As the last factor of effectiveness, grouping was used in most of the idioms. Naturally, the network flight map (idiom 1) represents the clearest example, with interactive filters permitting to isolate certain hubs to thus show all edges (flights) between selected cities. As an example, the screenshot below shows how the idiom can be manipulated to show only destinations reachable from Anchorage (ANC). Notice the origin airport is always displayed to the left of the route label.



Route Network Map Idiom - Grouping Destinations from ANC for Alaska Airlines

A second example of grouping is the idiom "Average Total Delays for Hubs by Airline" which, by means of a filter, includes delays only for airport hubs by airline. The matrix color code reflects the degree of delays, with darker shades of red indicating higher average delays for airport/airline combinations.

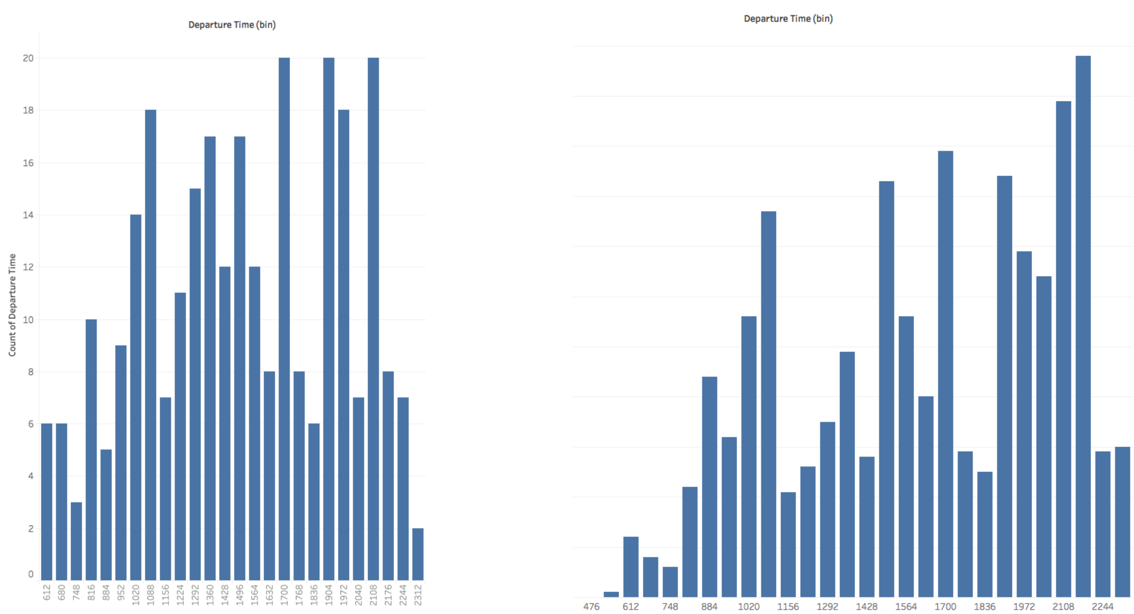


Average Total Delays for Hubs by Airline - Grouping Hub Airport Delays by Airline

In summary, a wide variety of channels is used across all 23 idioms. In each, the principles of expressiveness and effectiveness are used to improve the quality and degree of intuitiveness for each accomplished task.

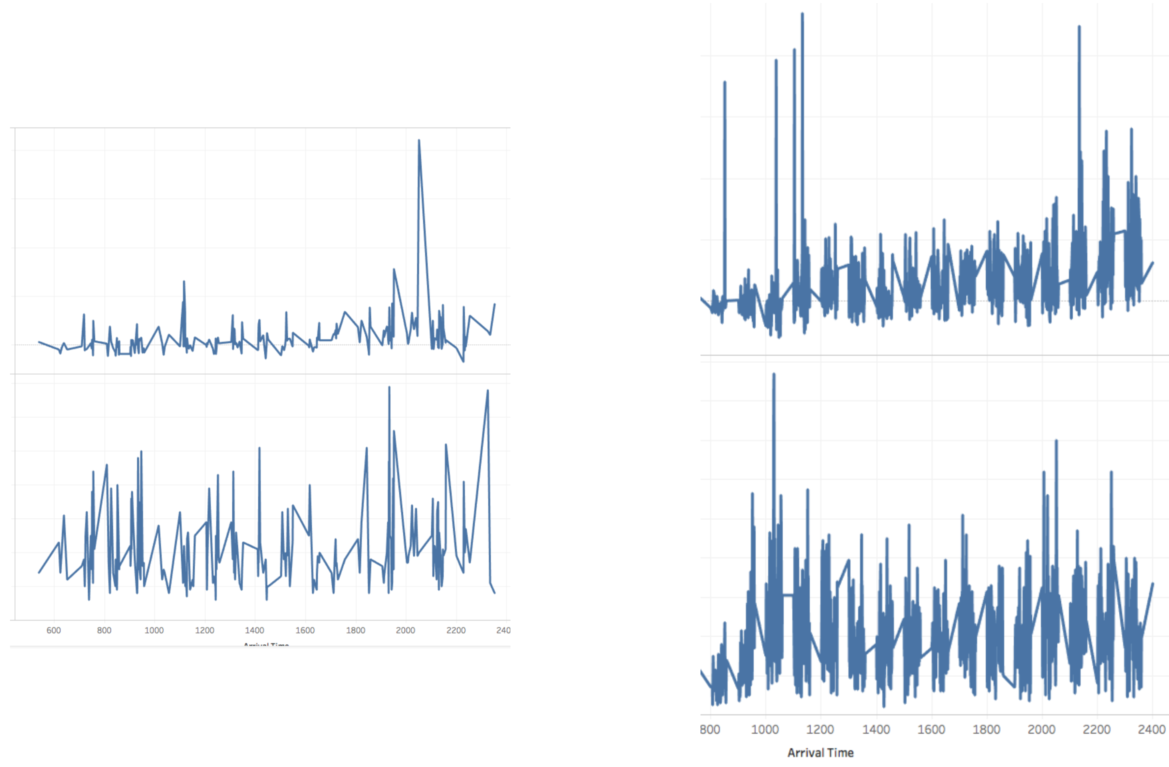
Key Findings:

Hub airport traffic is very different from non hub airports. This shown by idioms “SAN Departure Times through Day” and “ATL Departure Times through Day”. In the former, flight departures are more uniformly distributed through the day, while in the latter the hub airport displays the typical pattern of three large waves of departures (this is typical for hubs).



Left: Departures for SAN - Right: Departures for ATL (ATL has “three wave” pattern)

In addition, comparison between the “Non Hub Airport Arrival and Taxi In Delays” with “Hub Airport Arrival and Taxi In Delays” idioms shows that hub airport delays are much more highly correlated with taxi in times - they move hand in hand. This is a key finding which leads us to the last map of the story.



Left: "Non Hub Airport Arrival and Taxi In Delays" - Right: "Hub Airport Arrival and Taxi In Delays" (Higher degree of correlation on the right)

Difficulties, Hurdles, Weaknesses and Lessons Learned:

Finding ways to effectively plot all network routes on the same map was one of the main concerns. Unfortunately this project had to rely on only five days of the year as more data has caused software issues. A lot of informative trends have probably been left out as a result. In addition, this project could have relied more on statistical measures such as correlations and distributions to build a more rigorous solution. Finally, one last weakness is the use of Tableau: a programmatic solution would have allowed to perform dimensionality reduction techniques which would have yielded insights sooner in the project.

As a result of the research related to research papers, one key lesson from this has been that although more advanced visualization techniques may require time and effort to build, they are likely to yield more informative insights that would otherwise be learned through the construction of a wide variety of idioms and/or data exploration.

Bibliography

https://www.faa.gov/air_traffic/by_the_numbers/

https://www.rita.dot.gov/bts/press_releases/bts015_16

https://www.transtats.bts.gov/TRAFFIC/

http://www.aci.aero/

https://www.bts.dot.gov/explore-topics-and-geography/topics/airline-time-performance-and-causes-flight-delays

https://www.transtats.bts.gov/HomeDrillChart.asp

Research Paper 1:
Enhancing Visual Analysis of Network Traffic Using a Knowledge Representation
Ling Xiao* John Gerth Pat Hanrahan
Stanford University

Metric-Based Network Exploration and Multiscale Scatterplot
Yves Chiricota - Universite' du Quebec a Chicoutimi, Canada
Fabien Jourdan and Guy Melancon - LIRMM UMR CNRS 5506, Montpellier, France