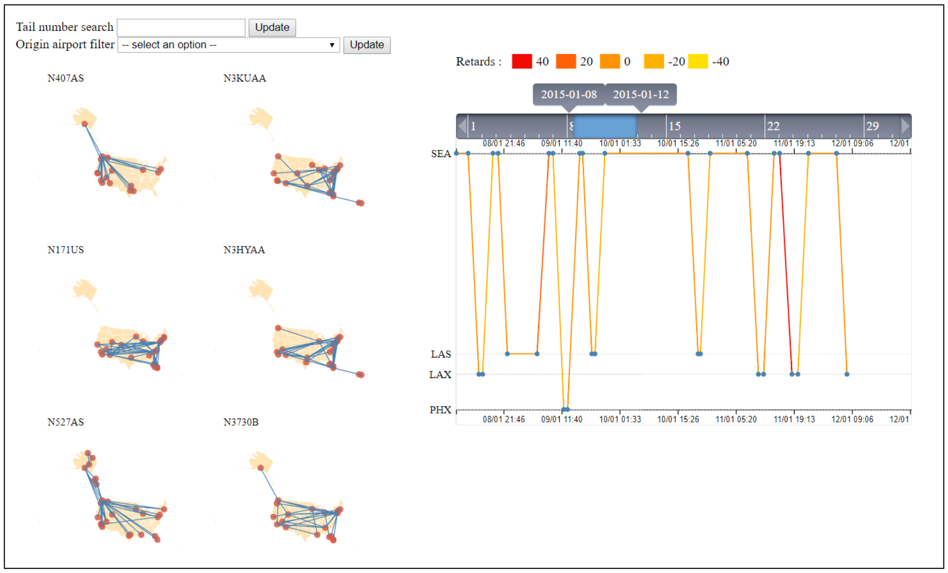
Interactive visualisation of flights in the United States at single-plane resolution

Arthur Aubret* Université Lyon 1
Caroline Gaud† Université Lyon 1
Brice Letcher‡ Université Lyon 1



Abstract

We propose a novel approach for visualising flight trajectories by taking a 'plane's eye view', allowing the detection of patterns in single-plane movements. We include search and filter functionalities for choosing the planes to visualise. Geographical mini-maps allow for a coarse view of flight patterns and for selecting the planes to focus on. For visualising precise trajectories, we revisit the Marey train schedule visualisation, showing the changes through time of flight distance and time, stationary periods and distance from airport of origin.

1 Introduction

Air traffic data is dense, multidimensional, and constantly evolving. It it hard to understand using only tables. Data visualisation techniques can play a key role in understanding the main trends in the data.

While we had an initial preference for analysing French air traffic data, and in particular departures/arrivals at Lyon Saint-Exupéry, open access data on flights is remarkably hard to obtain. We speculate this is for mainly commercial and security reasons. We thus concentrate on the best data we found, which is from the US.

In the United States (US), there are almost 5000 public airports. Over 9.5 million flights were recorded in 2015 carrying more than 900 million passengers [4]. The volume of traffic is therefore considerable.

There exist a wide variety of flight route visualisations based on maps, showing in many cases the global trends in air traffic. These reflect on the connected hubs of the network.

If the data includes unique aircraft identifiers for each flight, it is also possible to visualise air traffic patterns at single-plane resolution. We propose such a visualisation using aircraft tail numbers, which are the unique identifiers painted on the tail.

With flight data that includes tail numbers, it becomes possible to apprehend common flight routines for individual aircrafts. Airport bases for single planes can also become apparent, in terms of both flight connections and possible maintenance periods. Some insights might also be drawn from taking a single-plane view in terms of economic efficiency for airline carriers for example.

Flight pattern visualisations can be useful for airports and airlines to efficiently predict flights and better manage air traffic and air space. Their concern is to satisfy the largest number of clients while limiting the number of flights and the number of round trips. They also need to maximise plane availability to reduce costs.

To summarise, our main goal in this project is to visualise flight patterns in the United States at a single-plane level. We focus on showing spatial patterns and individual timed schedules using two complementary representations, which are geographical mini-maps and the Marey map.

Our visualisation is built using the D3 JavaScript library and presented on our GitHub web page [1].

2 Related Work

This section will describe the state of the art of flight visualizations in space and time.

2.1 Global trends visualisation

Most visualisations that represent air traffic show global trends using a geographical map. Figure 1 below, built Sophie Engle [5] using D3.js, is one such example.

A visualisation which brings a single-plane focus and is therefore closer to ours conceptually is shown in Figure 2 [8].

*e-mail: arthur.aubret@etu.univ-lyon1.fr
†e-mail: caroline.gaud@etu.univ-lyon1.fr
‡e-mail: brice.letcher@etu.univ-lyon1.fr

Figure 1: Main flight routes in the US using force-directed bundled

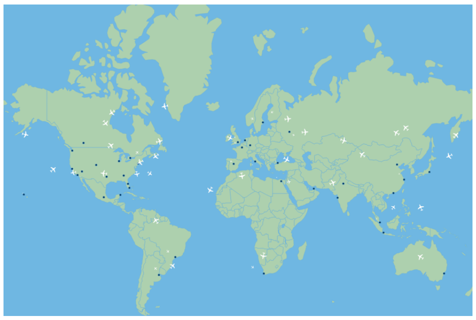


Figure 2: Animated single-planes on the world map in D3

This visualisation is an animation which gives the viewer a feel for flight density and explicit information on single plane origin and destination. The crucial information that is not available is trajectory, ie the succession of flights served by the same plane. This information is not amenable to being shown on geographical maps. For example, a colour coding on planes would bring visual clutter due to high numbers of planes and same connections across planes.

2.2 Work on Graphs

Since one of our representations is a geographical map including a graph, considering airports as nodes and flight routes as edges, we paid some attention on work related on graphs.

There is also work on temporality [3], which considers dynamic graphs and shows the evolution of a variable on an edge. One of the techniques uses an efficient image-based bundling method to create smoothly changing bundles.

Finally, many visualisations [7] and tools about air-traffic exist, but a significant fraction of them have different objectives from ours. For example, they are focused on accurately analyzing the trajectories of aircrafts.

2.3 Marey maps

Given we develop the visualisations in D3, we found particular appeal for one type of visualisations presented by D3 founder Mike Bostock: the Marey train map. The original Marey train visualisation, by E.J. Marey in 1885, is presented in Figure 3. The D3 implementation by Bostock is shown in Figure 4.

This visualisation is clearly adapted to our aim of showing single-plane trajectories in time and space. The lines between nodes can encode information on distance, time, both through velocity, or delay. We adapt the visualisation to the specifics of flight data relative to train data.

3 Project Description

3.1 Data

The dataset is taken from the open data platform Kaggle [6]. Over 5,800,000 flights over the single year of 2015 in the United States are recorded.

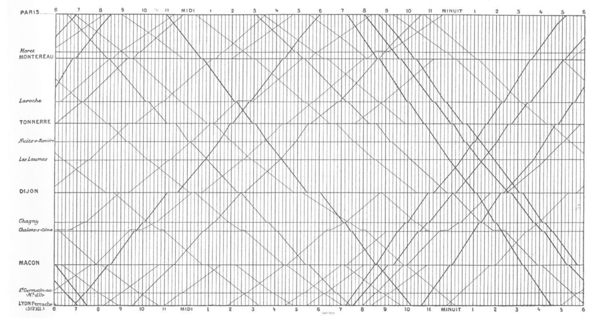


Figure 3: The original Marey train map

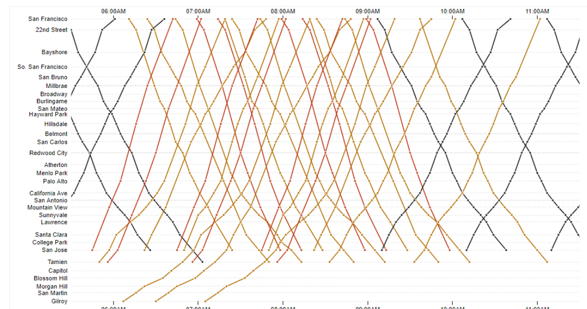


Figure 4: Mike Bostock's D3 Marey train map

The data was parsed to obtain only the information of interest including tail number, airport of origin, destination airport, departure time, flight time, arrival delay and distance flown.

Because of the large volume of the data, we focus on the month of January 2015. Grouping all flights with the same tail number together, we get 4399 unique flight trajectories of single planes.

The US airports, alongside their latitude/longitude coordinates are obtained from Mike Bostock's Git [2]. A total of 3379 airports are listed. We manually added a few airports included in the flights data but missing from this list.

We define an origin airport for each tail number as the maximum degree airport for that tail number: ie, the airport for which the plane passes through the most times. There are 42 such airports for January 2015. This highlights the small number of airports with high centrality.

3.2 Core concepts

Our building block is the dataset of flight routines of each aircraft in the United States, over the year 2015. Each flight routine can be represented naturally as a node-link graph drawn over a geographical map of the US. The nodes are airports, and the links flights between them. We can show many such miniature maps in a flight pattern panel.

This gives an among-plane overview of flight routines. In a complementary visualisation, we want to be able to generate the detailed ordered flight schedule for any aircraft. This gives the precise temporal trajectory of the aircraft from one airport to the next.

The minimaps should not encode too much information for risk of becoming cluttered and unclear. Because they will occupy a small amount of space, they cannot show the precise trajectory from airport to airport. This means the routines are shown as undirected rather than directed graphs.

The Marey map as defined for trains required adapting to flights data. Notably, a general trend is that planes make many connections, and often through the same airports. On this basis we define an 'origin' airport as the airport to and from which the plane most often flies.

This means making y-axis steps proportional to distances- as in Marey train maps- problematic if we are to show the whole itinerary. This is because showing the whole itinerary is only possible without duplicating airports on the y-axis: y-axis graduation is proportional to distance from the origin.

Otherwise, the frequently flown airports would have had to appear many times on the y-axis, making it unreadable. Our solution is to use line bends to encode distances between airports which might be at similar distances from the plane origin but at large or small distances between each other. We also provide a tooltip over those lines to display the distance, and highlight the corresponding flight in the minimap.

3.3 Interactivity

Interactivity is a central part of the visualisation. It allows the user to engage with the data and its representation directly.

The most important user interaction is clicking on a minimap of interest. This displays the corresponding trajectory in the Marey map, providing the detailed information lacking in the mini map. Selecting different mini-maps allows for trajectory comparisons.

The time scale on which the trajectory is shown is extensible. It can be enlarged and shrinked using a slider in the Marey map. The user can navigate through the days of the month by scrolling the slider.

Tooltips provide the user with airport names, flight distances, and times at departures and arrivals. Flight highlighting on the selected mini-map is provided when hovering above a flight in the Marey map, creating a visual connection between them.

3.4 The visualisation

The visualisation we created allows the user to answer a number of questions concerning air traffic in the US. The spatial scale is that of all US states including extra-continental Alaska, Hawaii and Puerto Rico, and the temporal scale is January 2015. The temporal scale is January 2015 but extensible to any month in 2015.

The questions that can be investigated include: How are plane itineraries organised? Do planes always come back to the same starting point, and over what temporal scale? Can we uncover spatial and temporal trends in single-plane trajectories?

Combining mini-maps and the Marey map offers two complementary views, not restricted to just picturing a graph over a geographical map.

The mini-maps provide the user with a global, qualitative, spatial feel for single-plane movements. Trajectory breadth and variey, and the number of served destinations, can be apprehended quickly.

Instinctively, the mini-maps allow the detection of broad or coarse-grained patterns. In Figure 5, we observe four planes among the set with origin the airport of San Francisco. We see that the four planes only serve close destinations along the SouthWestern Coast. For more details concerning the trajectories, such as whether the plane returns to San Francisco after each flight, we need the Marey map.

As a complement, the Marey map (shown in Figure 6) allows the user to gain detailed information on a pattern of interest. With a single click on a mini-map, this visualisation shows the detailed aspects of the data, ie the trajectories and distances through time. The visualisation displays the visited airports on the y-axis and time on the x-axis. The y-axis shows each visited airport only once and the distances are proportional to the distance from the origin airport.

A common issue arising in travels across the USA is to handle the different time zones. To do so, we computed all time in the time zone of one city, the first in our processed flights dataset.

A key innovation relative to Marey maps is the introduction of curved lines, in an attempt to solve the problem of distances on

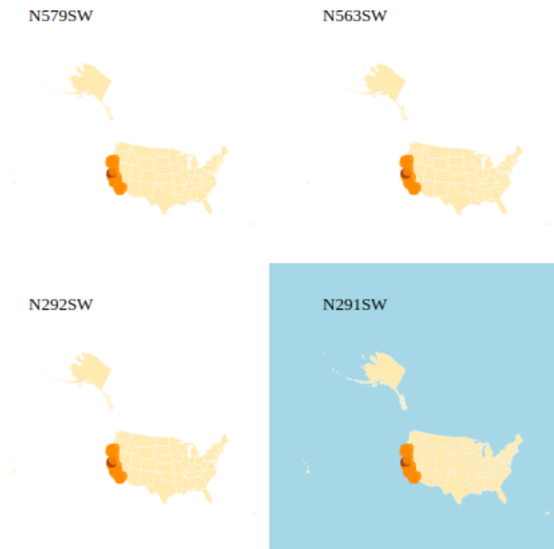


Figure 5: Four selectable minimaps. The nodes (airports) are highly clustered around the origin airport of San Francisco.

the y-axis (discussed in Core concepts section). The line total size gives the distance flown, irrespective of the distance to the airport of origin.

In this context the visualisation provides a vectorial indicator: we encode direction of movement (towards or away from the origin airport) plus distance flown.

Line colours notwithstanding (which adds another dimension), three key dimensions of the data are exposed in the Marey map: the two spatial dimensions of flight trajectory plus the temporal dimension. They are fitted inside the two dimensions of the Marey map.

Figure 6: Curved lines encoding travel distance.

In Figure 6, we see high curvature in the small space between Florida and Philadelphia. Charlotte, the city of origin, is actually located in between those two states. The plane has thus travelled a large distance without getting closer or farther to its origin airport

(Charlotte). High curvature can indicate a questionable trajectory in terms of maintaining high distance with airport of origin. It can also however question the legitimacy of our concept of origin airport.

A sliding window allows for moving across different days and its size can be modified to view larger or smaller time chunks. It is also possible to navigate inside the window by clicking left to go back in time and right to go forwards.

Upon hovering above a circle (an airport of departure or arrival) the departure/arrival time is displayed as a tooltip. Similarly, hovering above a link connecting two circles the number of kilometers and time taken for the correponding flight is displayed. Incidental information on flight delays is also provided, using a colour code on the links which gives the number of minutes before or after schedule.

3.4.1 Searching and Filtering

When first loaded, the visualisation shows a set of 10 mini-maps corresponding to the first 10 tail numbers in the data. The Marey map shows the trajectory of the first of those maps. Filters allow for access of any tail number, with autocompletion for the user unfamiliar with any of them, and focusing on specific airports.

Each time new tail numbers are chosen, 10 mini-maps are generated. This choice is motivated by the need for the mini-maps to be proximal to the Marey map so that the user can compare mini-maps and the Marey map directly and make use of the interactivity. One development would be to make the number of mini-maps displayed a modifiable parameter. Perhaps an efficient way of fitting more mini-maps into the view could also be found.

We implemented a search and a filter to give the user the ability to access the whole data and to bring out patterns in the data.

They are used as follows:

1. Tail number search: A specific tail number can be looked up using a search bar with autocompletion. This will show the corresponding minimap (and others) and Marey map.

2. Origin airport filtering: An airport of origin can be selected from a drop-down menu, which will generate a set of minimaps for which the flight itinerary has the largest number of flights to/from this airport. Pattern differences can appear.

4 Discussion

Below are a few examples of insights that can come out of our visualisation (we encourage the reader to check for him/herself and explore further!):

- We see that tail number N925SW frequently transits through Denver airport. It serves as a connexion between the MidEastern states and the West Coast. This observation, already instinctively visible through the mini-maps, is confirmed using the Marey map. When the plane is outside of Denver, it has many curved lines, which show that the plane is staying in the same zones. Denver thus acts as a bridge between Eastern and Western clusters.

- We observe some planes mediate mainly proximity flights (flights over short distances): they do not travel far and make frequent round trips between one main city and neighbouring cities. We see this for tail number N435SW centered in Los Angeles for example. The result is visible on the mini-map and confirmed in the Marey map.

- Some airports and planes serve as bridges between the mainland airports and the outer airports (Hawaii, Puerto Rico, Alaska). This is manifested as long repeated straight line patterns. As an example, we observe this pattern for tail number N547US between Los Angeles and Hawaii.

- We find some planes ensure one city's contact with the rest of the country. We observe for example that tail number N853AA connects Miami to the West Coast. A single intermediary is visible in the mini-map. The Marey map shows this recurring stop-over on the way to the West Coast is Dallas airport.

We thus emphasise the key role played by single plane flight routines in air traffic and for its airport of origin.

It is not however straightforward to observe a distinctive pattern for all tail numbers. The reason is that most planes actually combine several sub routines. We thus observe frequent transitions between different smaller patterns for single planes.

For instance, a plane might travel to the West Coast to make several small connections, then fly all the way back to the East Coast and start making long North-South connections. This could be the result of adjusting single plane routines to best meet flight demand. Conversely it could speak for poor trajectory optimisation and thus issues with trajectory management on behalf of the owning airline.

The nature of our visualisation is strongly influenced by our choice to assign each tail numbers an airport of origin. As a consequence, we observe a number of planes that travel many times to and from several different airports, showing that they can in fact have several 'bases' or origins. The Marey map visualisation is harder to interpret for those cases.

Our visualisation is nonetheless effective and we envision a generalisation to other datasets. It would for example be interesting to study international rather than domestic flights. This should highlight countries of origin more strongly than airports of origin at a national level. More generally still, our visualisation could be extended to single-unit maritime transport, road traffic, or any such means of transportation of goods or persons.

5 Conclusion

The proposed visualisation provides the user with a wealth of information without generating visual clutter. Clutter would be inherent to a geographical map-only representation, while the Marey map avoids this problem. We make use of a variety of encodings, including shape, colour and distance, to highlight features in the data with the intention of keeping with visual clarity for the user.

To our knowledge, no visualisations represent single-plane trajectories in time and space on flights data. We thus provide a complementary and innovative alternative to the common global trends representation of flights data.

Our visualisation proves effective in allowing multi-level (time, space) and multi-scale (coarse with mini-maps and fine with Marey map) analysis of flight trajectories. Trends turn out to be uncoverable despite the initial lack of coherence in flights data. The complexity of air traffic data is also apparent at single-plane scale, where sub-routines and multiple airports of origin are frequent. The visualisation is generalisable to any single-unit transportation trajectories.

Several important improvements are possible. It would for example be interesting to extend the time frame selection slider on the Marey map to the mini-maps so that the points shown are the same between the two.

Along the same lines and for a more flexible visualisation, we could add mini-map rearrangement and enlarge the selected mini-map with a modifiable zoom. This allows for apprehending geography more directly. It is currently difficult to compare the Marey map and the lowermost mini-maps in the pane, requiring a search of the tail number of interest. Similarly, making the number of displayed mini-maps specifiable would add flexibility.

We would like to address special thanks to Romain Vuillemot for providing the idea for this visualisation.

References

[1] C. G. Arthur Aubret, Brice Letcher. Interactive visualisation in time and space of flights in the united states at single-plane resolution. https://datavisuproject.github.io/US_Flights/.

[2] M. Bostock. Airports. https://gist.githubusercontent.com/mbostock/7608400/raw/airports.csv.

[3] S. I. F. T. R. K. A. C. T. Christophe Hurter, Ozan Ersoy. Bundled visualization of dynamic graph and trail data. *IEEE*, 2007.

[4] U. S. department of Transportation Office of Public Affairs. 2015 u.s.-based airline traffic data, 2015. https://www.rita.dot.gov/bts/press_releases/bts018_16.

[5] S. Engle. Flight paths edge bundling. 2017. https://bl.ocks.org/sjengle/2e58e83685f6d854aa40c7bc546aeb24.

[6] F. Jonsson. Flights. 2017. https://www.kaggle.com/freddejn/flights2.

[7] M. D. Ronald Azuma, Howard Neely III and R. Geiss. Visualization tools for free flight air-traffic management. *IEEE*, 20:32–36, 2000.

[8] Tnoda.com. Flight animation with d3.js. 2014. http://www.tnoda.com/blog/2014-04-02.