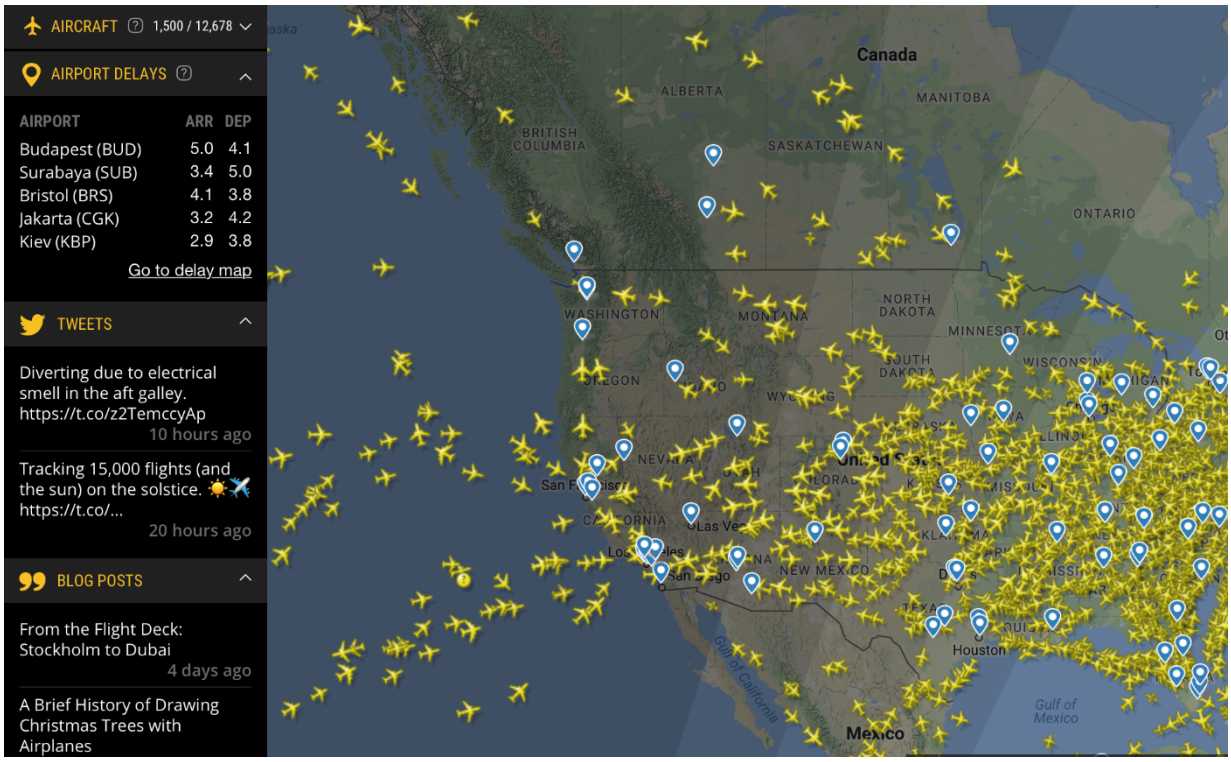
US FLIGHT

Data visualization – Process book – fall 2017

Kévin Brou Boni, Clément Pavué, Mohamed EL Mehdi Roussaky

Overview, motivation and target audience

Try to google “Flight tracker" on the internet, you will see a bunch of website with several features. As an amateur, the visualization could be painful. On the one hand, we have beautiful visualization, but on the other hand there is practical visualization. An example below from the website https://www.flightradar24.com/.



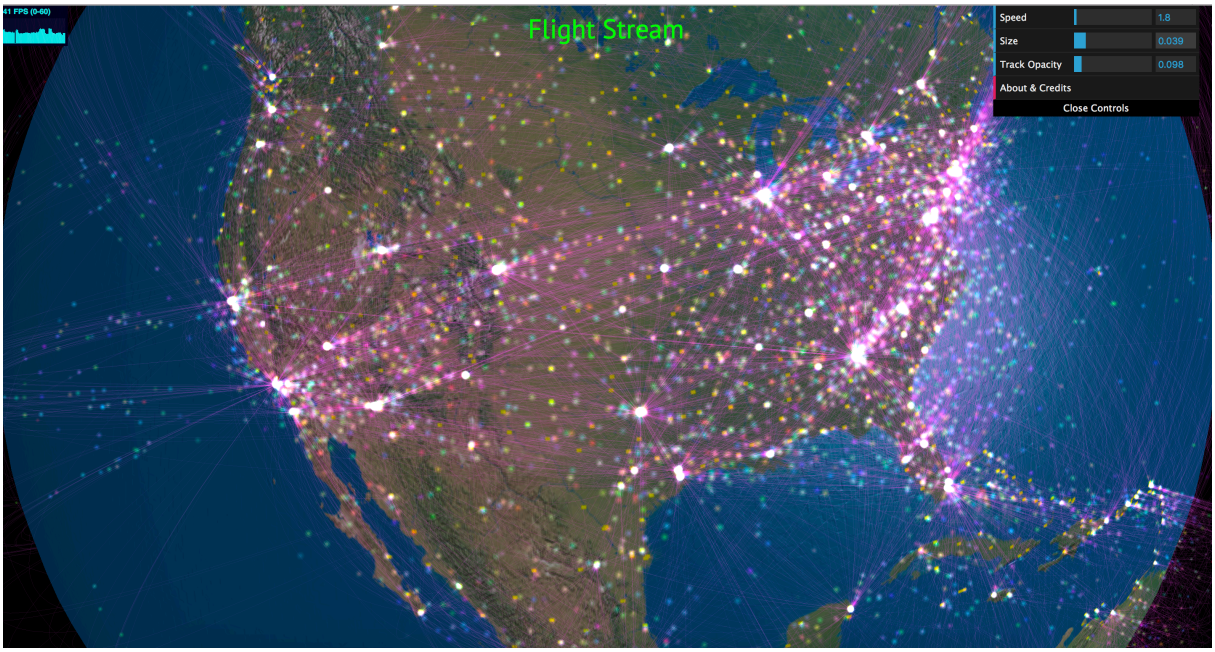
Most of them are identical, information is not easily accessible and readable. An obstacle course begins to understand, filter then select needed information. We could not criticize these platforms because they are targeting aviation enthusiasts.

In this project, we are willing to offer an accessible visualization of flights and airports. As a group of amateur, the idea is to quickly explain some facts associated with data visualizations.

The target audience is not only aviation amateur, it will be a platform where we first tell a story with data visualization. Therefore, all the curious greedy to learn some facts about aviation in general are also welcome.

Related work and inspiration

Our goal is to combine comfort and convenience. The first related work was seen in on the #help channel in slack:

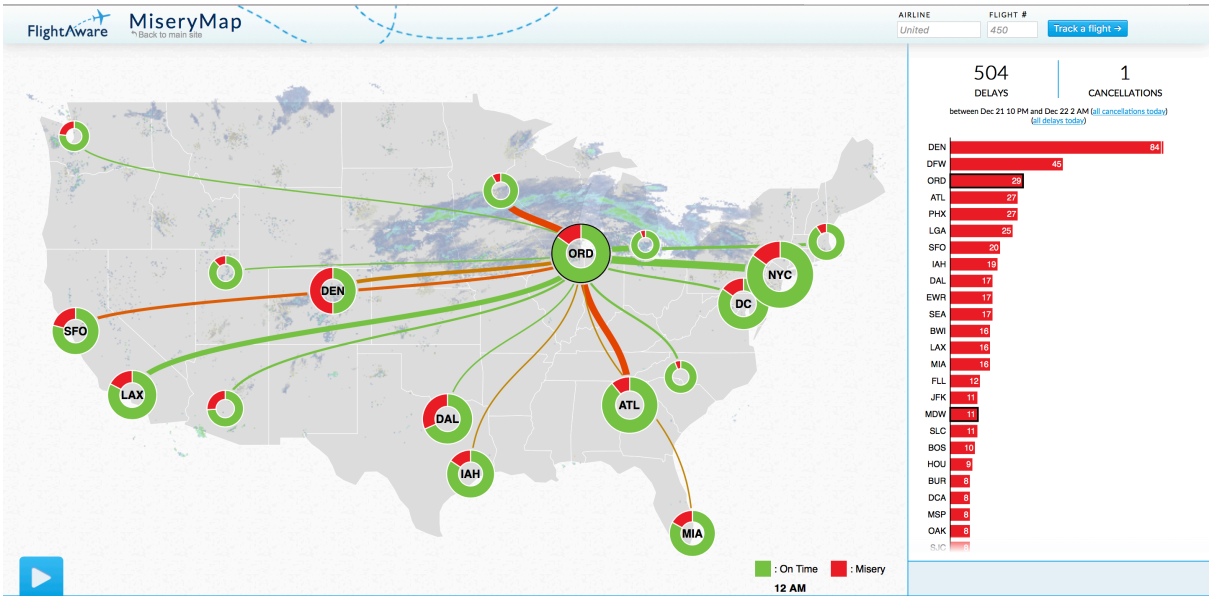


Source: callumprentice

What we liked about this visualization is that the user has an interface in order to control the flight traffic (speed, size). That is beautiful and enjoyable but not very useful. There is no legend or other information except several dots moving. We can easily understand that they represent an airplane nevertheless we don't any information on their routes.

This map is convenient to have an overview on the flight traffic but in order to find the weather or the current traffic airport it is pointless. We can face it by reducing the number of interactions. Lower the number of airports, combine different visualization tools to save space.

A second inspiration source shows the delays across major US airports:



Source: FlightAware

This visualization has less feature but guarantee a good understanding. Airports are represented by a dot with his code and a gauge showing the delays average. User can select an airport by clicking on it, then some continuous weighted lines representing an air road are displaying.

What we like here is the intuitive experience. No legends or numbers, this visualization explain itself. Moreover, a cloud shape is displaying, a manner to maybe justify delays.

Questions: What am I trying to show this my viz.?

Allow any person to know according to a city where he can travel. A convenient tool displaying the minimum information like the forecast, airlines available and their frequencies.

On top of that, giving tips about airlines like their cruising speeds or the delays in airports.

Our goal is to provide a minimalist tool, it is not a 24 hours live tracker. We want through the visualizations giving the whys and the wherefores of United States history in aviation

Dataset: where does it come from, what are you processing steps?

Our dataset comes from Kaggle, upload by Department of Transportation. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations.

The dataset contains 3 files, the first rows below.

- airlines.csv:

IATA_CODE	AIRLINE
UA	United Air Lines Inc.
AA	American Airlines Inc.
US	US Airways Inc.
F9	Frontier Airlines Inc.
B6	JetBlue Airways

- airports.csv:

IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
ABE	Lehigh Valley International Airport	Allentown	PA	USA	40.65236	-75.44040
ABI	Abilene Regional Airport	Abilene	TX	USA	32.41132	-99.68190
ABQ	Albuquerque International Sunport	Albuquerque	NM	USA	35.04022	-106.60919
ABR	Aberdeen Regional Airport	Aberdeen	SD	USA	45.44906	-98.42183
ABY	Southwest Georgia Regional Airport	Albany	GA	USA	31.53552	-84.19447

- airlines.csv:

YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE	FLIGHT_NUMBER	TAIL_NUMBER	ORIGIN_AIRPORT	DESTINATION_AIRPORT	SCHEDULI
2015	1	1	4	AS	98	N407AS	ANC	SEA	
2015	1	1	4	AA	2336	N3KUAA	LAX	PBI	
2015	1	1	4	US	840	N171US	SFO	CLT	
2015	1	1	4	AA	258	N3HYAA	LAX	MIA	
2015	1	1	4	AS	135	N527AS	SEA	ANC	

This dataset is clean, moreover there are much more features needed. Relatively new, we have still to reduce dramatically his size (more than 700MB). The preprocessing steps:

-**Explanatory analysis** (python) in order to estimate the possibility of reach our goal.

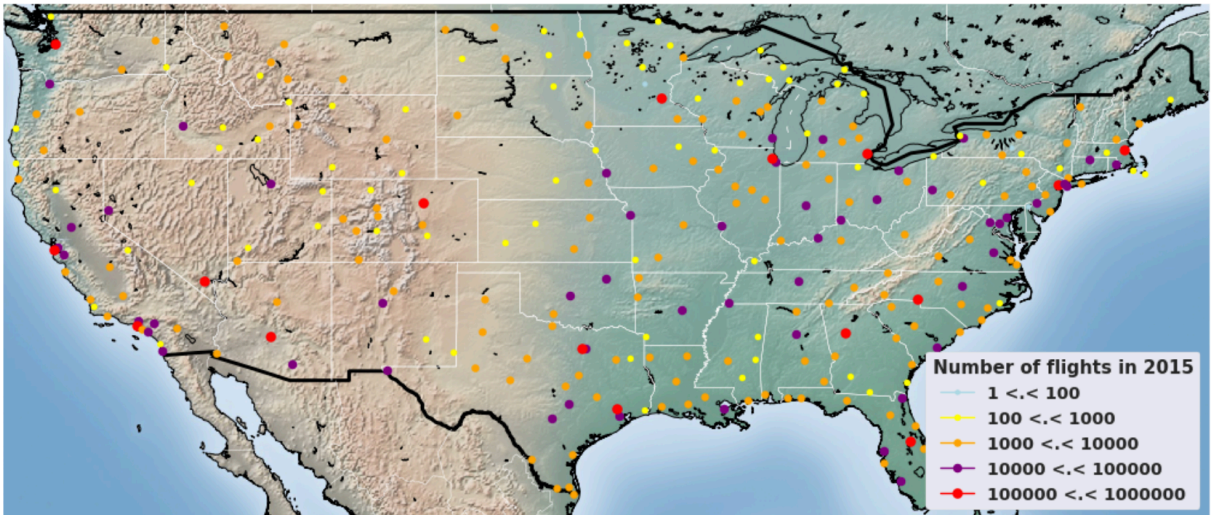
-**Reducing and splitting the data** (python) in smaller csv file. This step was mandatory in order to decrease the computing resource so time in the navigators.

-**Add some features** (python) like the weather thanks to the Dark Sky API.

Explanatory Analysis: What visualization have you used to gain insights on the data?

We dramatically decrease the computing time by only taking in account 2 months of data. Previously we verified that the distribution is followed.

First of all, we had to draw a US map showing the airports and their number of flights in 2015:



Jupyter Notebook

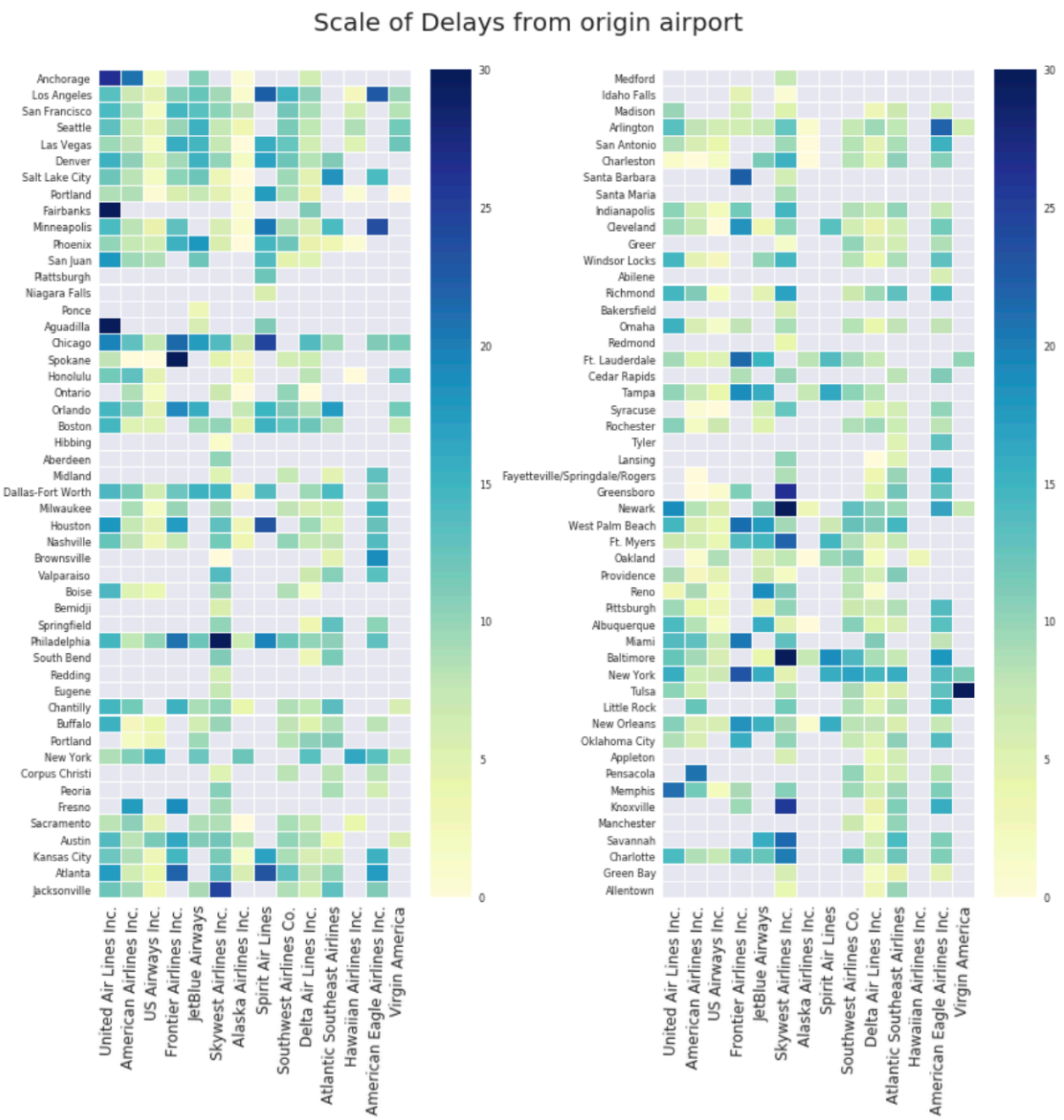
There is a huge number of flights in US. We decided to divide airport in 3 categories:

-L: more than 200 flights a day.

-M: between 200 and 50 flights a day.

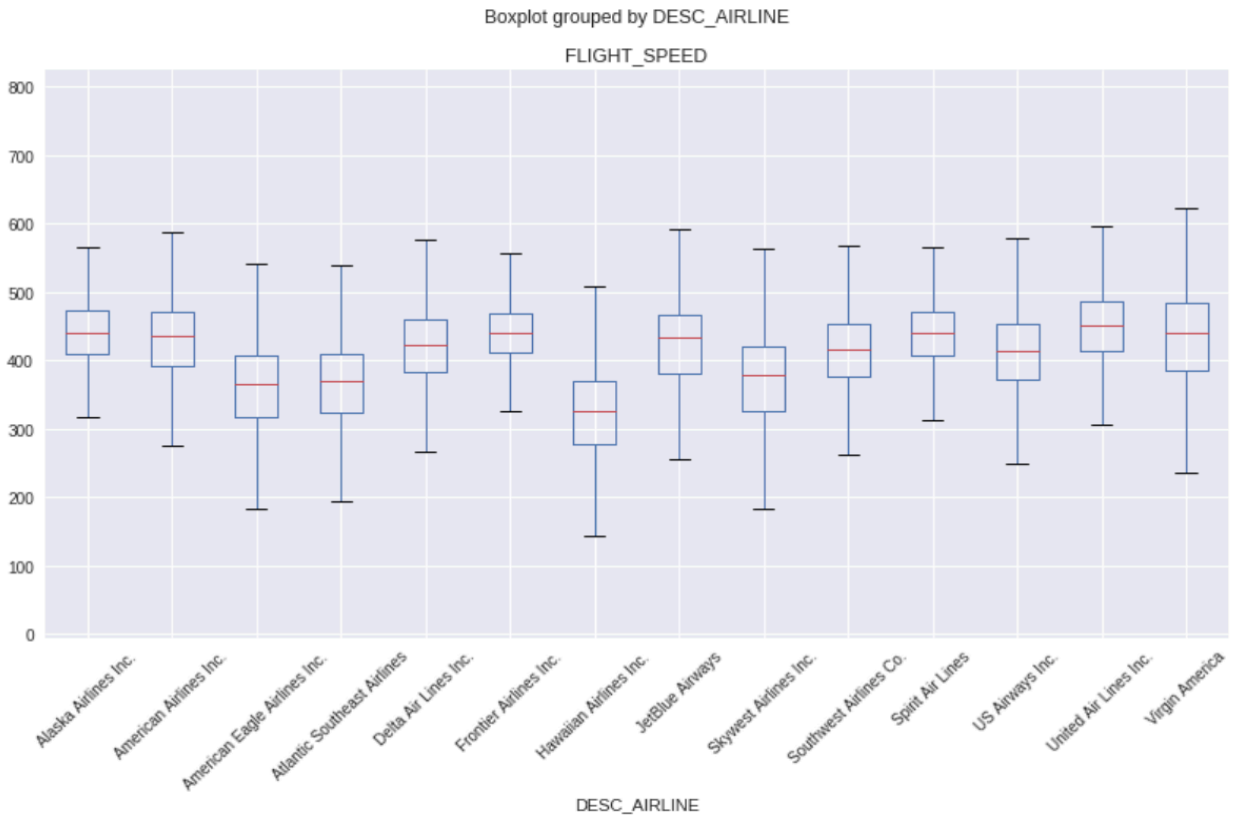
-S: less than 50 flights a day.

In a second part, we would know how airlines are responsible of delays. There were too many airports for a basic plot, the best decision was to plot a heat map.



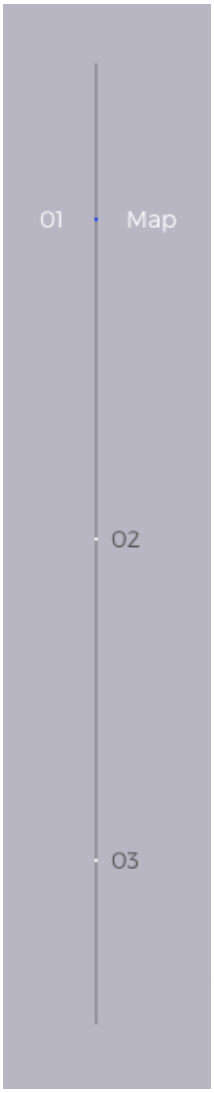
Jupyter Notebook

We raised that even with a departure delay some flights always arrived on time. The average speed is higher from a normal flight to make up for lost time. The idea of going deeper in the cruising speed came. Below a box plot of speed (mph) by airline.



Jupyter Notebook

Implementation:



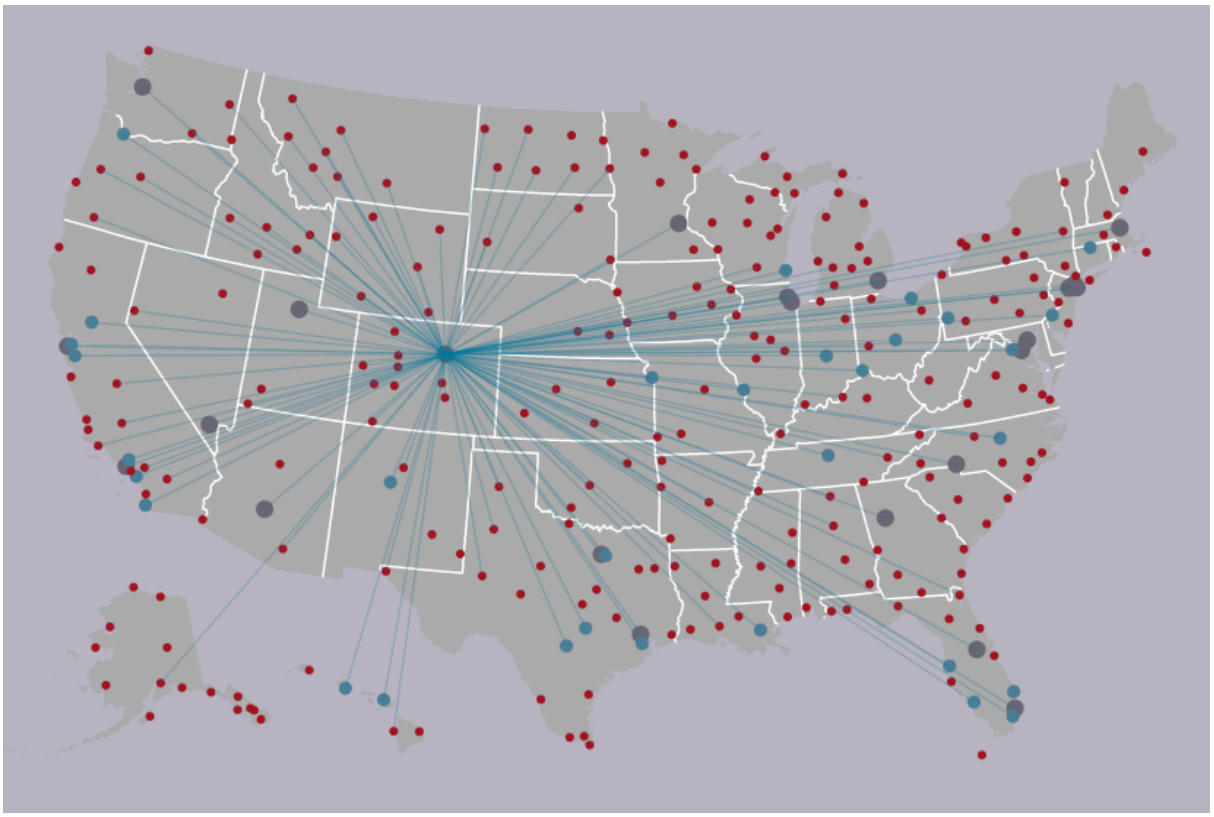
Our website is built like a book, there is chapter. You don't navigate by scrolling but by turning a page. It is a continuous flow, on each page you can read a small story before interacting with the visualization. You are guided through the process.

You can navigate either using the scroll bar on **(Fig. 1)** or by selecting in the menu on the top right corner **(Fig. 2)**.

Fig. 2: Menu navigation

Fig. 1: Scroll navigation

Fig. 3: Map



Each dot represents an airport. There are three types as explained above, click on one to display his information on the right side (**Fig. 4**).

Fig. 4: Airport information

Fig. 5: Legend

Click on one circle to filter and select only this type of airport. Click a second time to come back to the initial state.