# Visual Analytics - Project proposal

Vaerio Colitta, Davide Spallaccini

No Institute Given

**Résumé** We want to tackle the domain of flight across the US. The dataset has been take by the `Bureau of Transportation Statistics` website.

## 1 The dataset

The dataset contains all the flights that were made across the U.S. during 2017.
The size is huge..it contains roughly $500 \times 10^3 \times 12$ tuples, representing one flight each, so preproccessing is necessary to work with this data in real time.
We also want (for qualitative metrics), to extract for each month, 1.000 tuples, in order to plot them easily.

## 2 The views

We want to have several coordinated and not coordinate views, to have some insigths of the most important patterns of flights.
First, we'll show a map of the U.S., in which each state is colored according to a color-scale which describes how many flights have left or arrived in that particular state. This is done month-wise, so we have a slider to switch between months.
Whenever a month is switched, the map changes its color accordingly to the new values for each state. Together with it another view changes. This one is a bar chart, where we have all the U.S states on the x-axis, and the toal number of flights as the height. This is done in order to compare them side by side. Moreover, we want to show the portion of flights that have left the state, and the portion of the ones that have landed on it in that particular month.
Since comparison may be difficult, if a user clicks on a bar, a line on the x-axis is plotted to show the differences with the other States.

We decided to put another non-coordinated view that shows the behaviour of the flights for each month, in order to see overall what are the months which have more flights that others.

With the first three views, we don't get where these flights went to, we just know that they have landed/left the state. With that being said, we use a `Chord`

`Diagram` to show the relationships among the states. If a user clicks on a state, the chord diagram will show only the pairs (selected state - other states).

Finally since in our dataset we have also information on whether a flights has delayed, we would like to reduce the dimensionality and plot them in order to be able to find clusters.
The user can select two or more elements and see them in hight dimension, by the use of a parallel coordinates chart