Assignment 2

Guy Ackermans 1227983, Niels Drost 0777877

January 29, 2017

In this research, some of the visual analyses that are possible on a data set on commercial airline flights have been investigated. An interactive program has been made which features a globe that can be used by the user to select the departure and destination airports of a journey. All possible flight routes between these airports are subsequently displayed on the globe, as well as complementary information in the form of a line chart and a histogram. In this report, first some possible user questions are discussed. Then it is discussed what diagrams could best be used to assist the user in finding the data he is interested in. The next section discusses the details of the implementation of the application and the interaction between the implemented diagrams. Then, some data exploration examples that can be done using this application are given. Finally, this report concludes with some suggestions for future research.

1 Introduction

In this research, some of the visual analyses that are possible on a data set on commercial airline flights have been explored. Our goal was to create an application that can assist travellers in determining which flights are the best choice for them, and that can also be used by airports and airlines to optimise their services by finding bottlenecks in their services. We decided to focus on two things: comparing the average delays of flights, and comparing the airlines that provide flights. Our application has been made using the D3.js library[1].

The user of our application should first be able to provide his planned departure airport and his planned arrival airport. The user should then be able to quickly determine which flight option is the least probable to be delayed, and which flight options are most likely to be provided by his favourite airline. The user should also be able to discern whether his favorite airline does not have exceptionally high delays compared to other airlines regarding his planned flight. To give a clear overview on the user questions that the application should be able to answer, the following enumeration of questions is given:

Q1 To which international airports in the United States are there flights from airport X anywhere in the world?

Q2 What are the relative number of flights from airport X anywhere in the world towards all airports in the United States?

Q3 What are possible transfer airports between airport X anywhere in the world towards an airport Y in the United States?

Q4 What are the seasonal effects on average departure delays of the 20 largest airports in the United States?

Q5 What are the seasonal effects on average departure delays of the airports that are capable as a transfer airport between airport X around the world towards an airport Y in the United States?

Q6 What are the seasonal departure delays for the airlines flying from an airport Y in the United States?

Q7 What is the distribution of airlines for each airport flying from the 20 largest airports in the United States?

Q8 What is the distribution of airlines for each airport flying from an airport that is capable as a transfer airport between airport X around the world towards an airport Y in the United States?

Q9 What are the seasonal flight delays per airline for a specific airport?

2 Choice of Visualization Techniques

2.1 Flight Routes

In order to allow users to quickly and easily input their departure and arrival airports, and to compare flight routes and flight details with each other, we decided that the user should be able to visually see all of this information on his screen. Our application therefore requires an interactive globe on which flights and airports can be displayed. It should also be possible for the user to select his departure- and arrival airports on this globe. All flight options between these two airports can then be visualized for the user, and further information can be provided so the user can compare details of these flights.

2.2 Flight Delays

Since it is not possible to completely predict a flight delay beforehand, we want to provide the user with data on the delays of past flights. The user should be able to easily see, for each month of the year, which of his flight options have had the highest and lowest number of average minutes of delay in the past. A line chart is the best way to visualise this data: The month of the year can be displayed on the X-axis, the average departure delays in minutes on the Y-axis, and each flight option can be represented on the chart by a line, made distinguishable from each other with the use of labels.

2.3 Flight Providers

The decision making process of the user on which flight option is the best for him or her, can also be facilitated by showing the user for each of the flight options, which airlines usually provide these. Usually, a person who wants to travel with an airline of his choice, has to browse through each of the individual flights departing on that day, to see whether the flight is provided by the preferred airline. When there are many flights departing for his destination, this can become a very time consuming process. Our application will have to make this process easier for the user, by presenting a diagram that the user can use in order to quickly see which flight route is most likely to be provided by his favourite airline.

The best option to display this data is by histogram, with the airports on the X-axis and the number of provided flights on the Y-axis. Each bar on the histogram will be subdivided into several colors, each color will represent a different airline. We originally contemplated using a pie chart to display this data, but we abandoned this idea in favor of a histogram because subdivisions can be displayed more clearly in a histogram than in a pie chart.

3 Research Data

Several publicly accessible data sets have been used for this research. The names and locations of airports around the globe have been obtained from OpenFlights[2]. Data on international flights has been obtained from The United States Department of Transportation[4], however this data set did not contain many features. We therefore complemented our research data with another data set from the The United States Department of Transportation, on domestic flights within the USA[3]. This data set is rich on features, including the average minutes of delay of a flight for example.

In order to reduce the time complexity of the application, the research data has been filtered in such a way that only the 300 airports that provide the most flights towards airports within the USA remain. The application can be used to view all of the flights between these 300 airports around the globe. However, information on flight delays and airlines can only be seen when either the departure airport or the destination airport (or both) are located in the USA.

4 The Globe

Figure 1: The Globe.

The core of our application is the globe (see figure 1). Every dot on the globe represents an airport of our data set, the size of each dot represent the number of flights arriving and departing from that airport. The globe is fully interactive: The user can zoom and drag the globe with his mouse. Hovering the mouse over a dot will make a label appear with the name of the airport. The drop-down menu in the top right can be used to instantly navigate to an airport; the globe will automatically rotate to bring the selected airport in view.

The user can select two airports simultaneously: One by clicking with his left mouse button, the other by clicking with the right mouse button. The user should start by left-clicking on the airport from which he wishes to depart. This will highlight all of the airport's flights inbound for the U.S.A. on the globe (see figure 2).

Next, the user can select his destination airport by clicking its dot with the right mouse button. This will show all flights option between his airport of

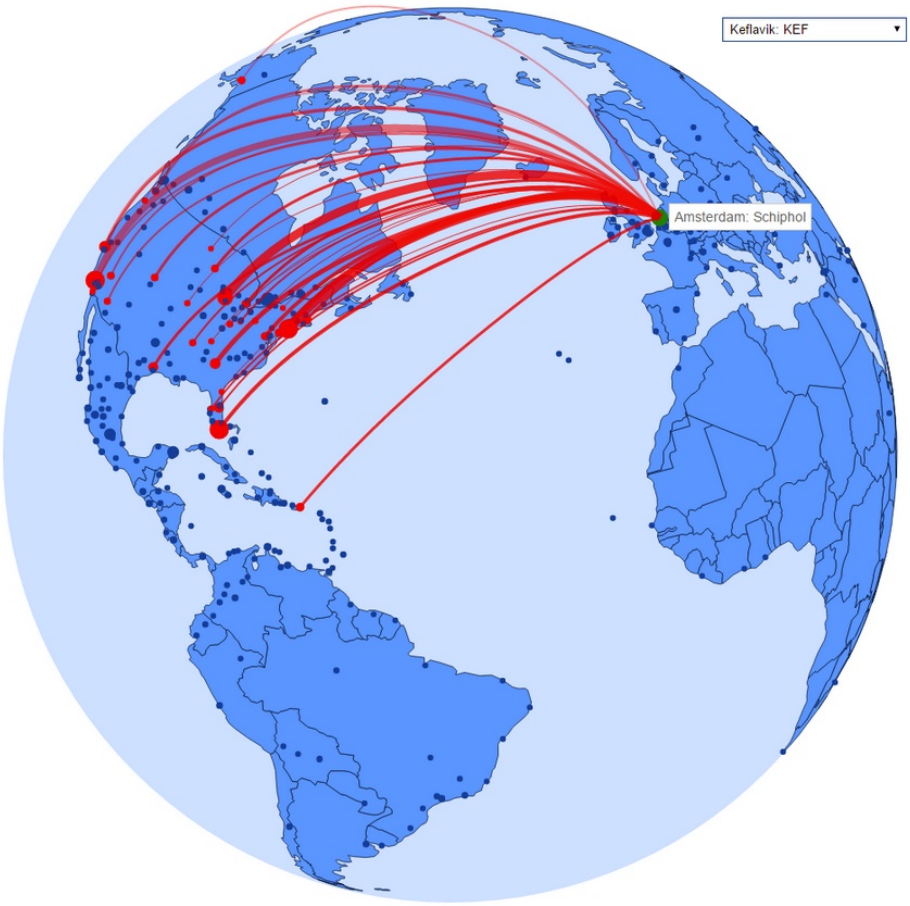

Figure 2: All flights that depart from Amsterdam: Schiphol with a destination in the U.S.A.

departure and his airport of destination, including indirect flights. The first flight of the journey will be represented with a red arc, the second flight with a green line. If applicable, the direct flight between the two airports is represented with a yellow arc. See figure 3 for an example of two airports selected with a direct flight between them, and figure 4 for an example of two airports selected without a direct flight between them. The globe can be used to answer Q1, Q2 and Q3.

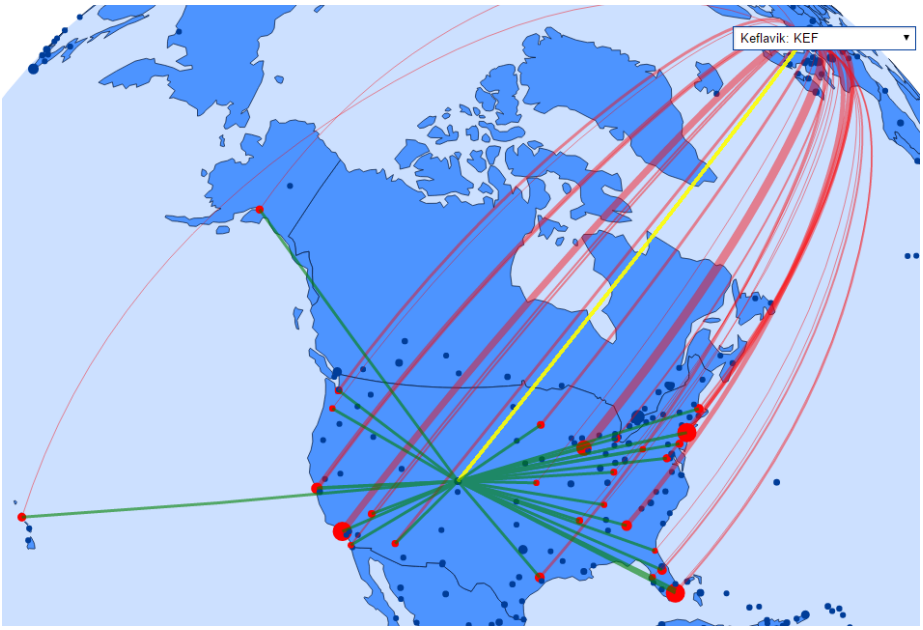

Figure 3: Two airports selected with a direct flight possibility between them. Departure airport: Schiphol, destination airport: Denver.

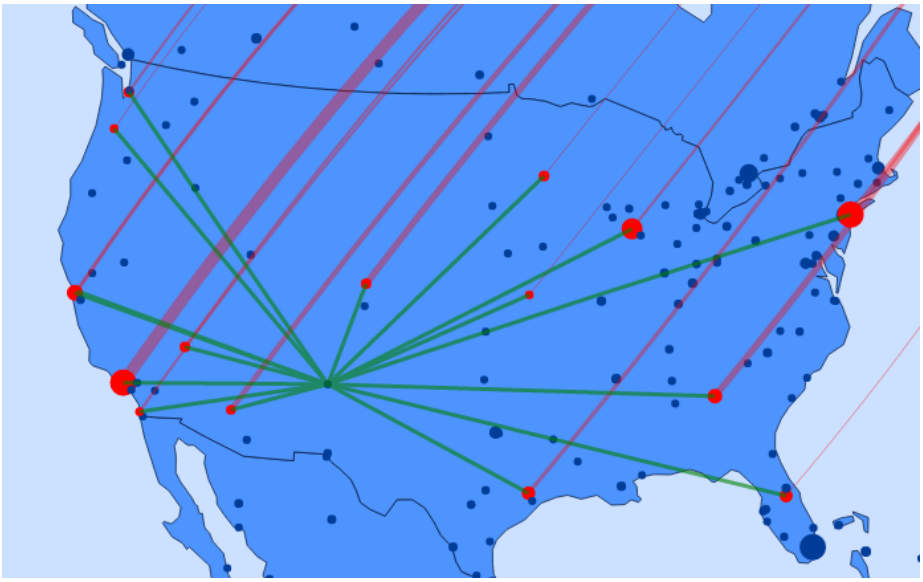

Figure 4: Two airports selected with only indirect flights between them. Departure airport: Schiphol, destination airport: Albuquerque International Sunport.

5 Flight Delays Line Chart

5.1 Delays per Airport

Once the user has selected his destination and arrival airports, as shown in figure 4, a corresponding line chart showing flight departure delays from the past will be displayed, as shown in figure 5. Light colors indicate naturally small delays, whereas darker colors indicate higher delays. The user will immediately know that if he wants to fly in July, it might be better not to fly via New York city if he wants to mitigate the risk of a flight delay as much as possible. In order to make it clearer which line of the line chart corresponds with which flight route on the globe, the user can hover with his mouse above a line. This will highlight the corresponding airport on the globe. Similarly, the user can also hover with the mouse above an airport displayed on the globe, this will highlight the corresponding line in the line chart. This feature can be used to answer Q4 and Q5.

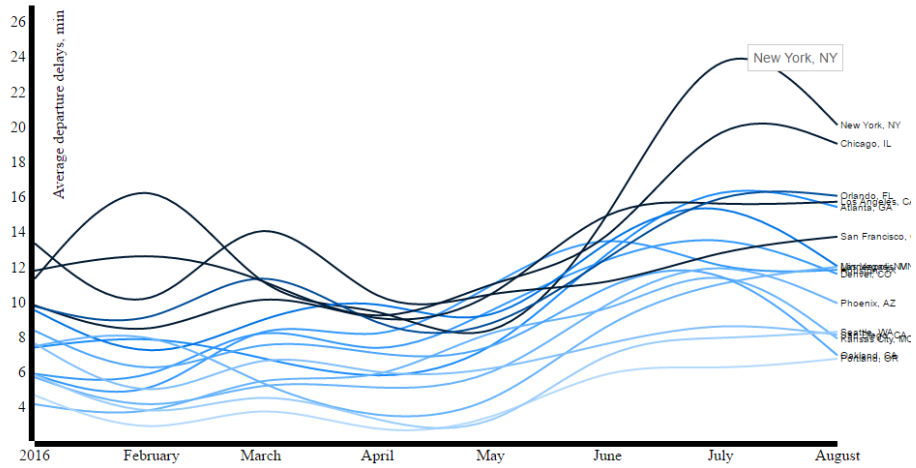

Figure 5: Line chart showing past flight delays. Departure airport: Schiphol, destination airport: Albuquerque International Sunport. The color gradient represents the total average delay of each airport: The darkest line has the highest average delay over the year.

5.2 Delays per Airline for a specific Airport

The lines that represent airports in the line chart discussed in section 5.1 can be clicked on to reveal additional information. Clicking an airport line will show the average delays of each airline on that specific airport. See figure 6 for an example. This functionality can be used to answer Q6.

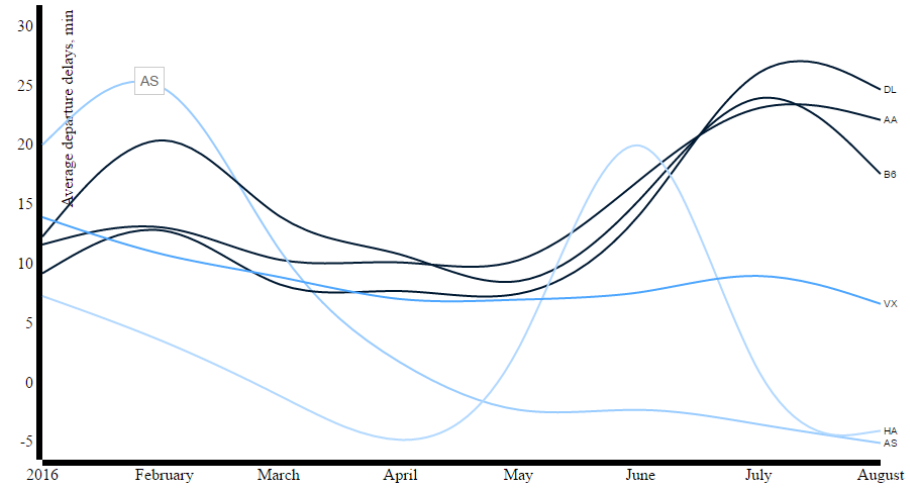

Figure 6: The new line chart after clicking on the "New York, NY" line in figure 5. Each line now represents an airline that provides flights from this airport to the destination airport.

6 Airlines Histogram

6.1 Airlines per Airport

Simultaneously with the line chart of flight delays, a histogram will be displayed once the users has selected his destination and arrival airports. The user can use this to quickly find which airport route on the globe he should take if he wants to fly with his airline of choice. if the user hovers with his mouse over the airlines legenda, the corresponding airline will be highlighted in the histogram. Similarly to the flight delays line chart, the user can hover his mouse above the histogram bars in order to light up the corresponding flight route on the globe. Hovering with the mouse above a flight route on the globe will also highlight the corresponding bar on the histogram. This feature can be used to answer Q8.

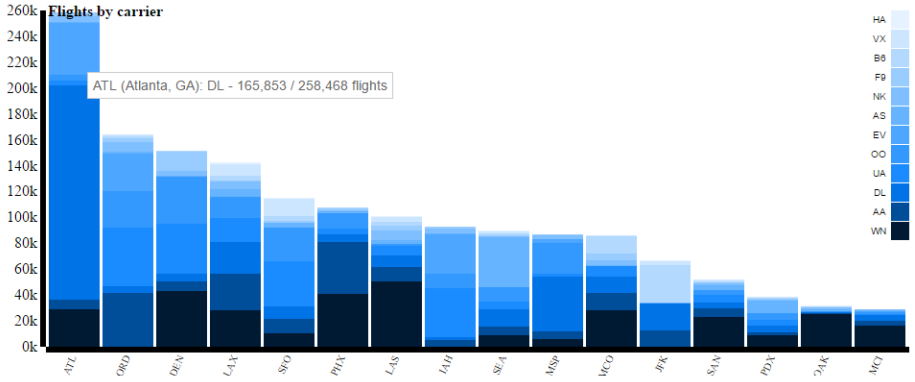

Figure 7: Histogram showing airlines per airport, and the number of flights provided by each airline. Departure airport: Schiphol, destination airport: Albuquerque International Sunport.

6.2 Interaction between Line Chart and Histogram

When the user has clicked on an airport line in the line chart in order to see the delays per airline for that specific airport (see section 5.2), additional information can be viewed in the histogram. The histogram discussed in section 6.1 will still be displayed, but when the user hovers with his mouse above an airline line in the line chart, the corresponding airline will be highlighted in the histogram. See figure 8 for an example. This feature is built in to answer question Q9.

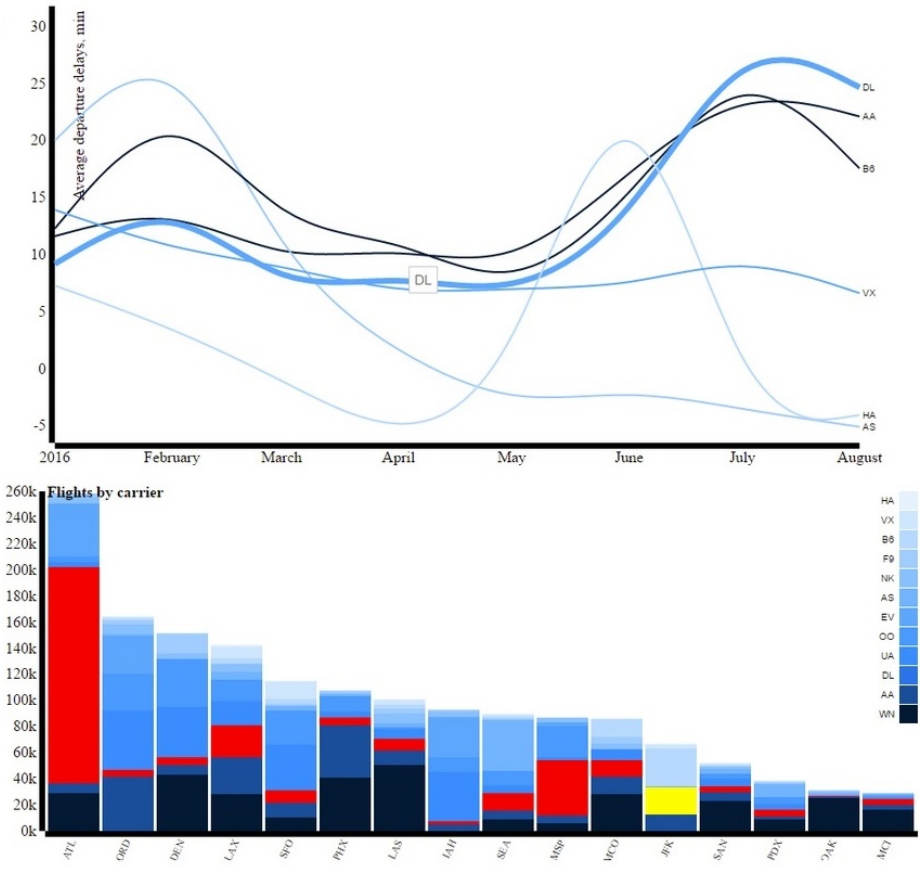

Figure 8: Interaction: Hovering the mouse above a line in the line chart will highlight the corresponding areas in the histogram. Departure airport: Schiphol, destination airport: Albuquerque International Sunport. The airport line which was clicked on: New York, NY. The mouse hovers above the line of the airline "DL" in the line chart. All occurrences of DL in the histogram are immediately highlighted with red, and highlighted with yellow in the bar that corresponds with the airport selected in the line chart.

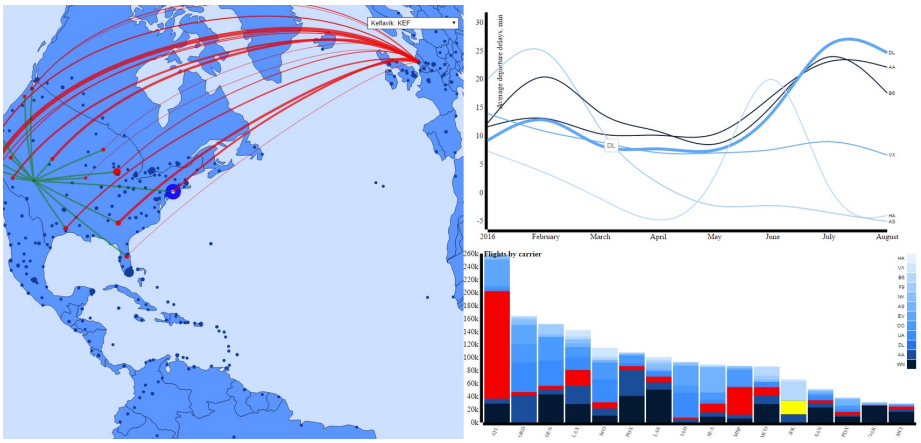

Figure 9: The application in its entirety. The selected airports are: Departure airport: Schiphol, destination airport: Albuquerque International Sunport. The airport line which was clicked on: New York, NY. Note the interaction between all 3 diagrams: The airport of New York City (JFK) has been highlighted on the globe by way of a blue circle. The line chart shows data on the average minutes of delay per airport, and hovering the mouse above the line representing the airline "DL" has resulted in the highlighting of all of the occurrences of DL in the histogram.

7 Data Exploration

In this section, a few examples of data exploration that can be done using our application are discussed. For instance, when we select a random airport, such as Moscow in Russia, only the largest airports in America are presented as destination airports. This example is shown in figure 10. Moscow does not fly to smaller cities in the US. Cities in Mexico on the other hand, do fly to small cities but almost entirely to cities in the south-west region of America, the Latin cities. A likely explanation for this is that Mexicans who work in these cities regularly want to visit their relatives back in Mexico by airplane. An example is shown in figure 11.

Airports within large cities, such as Los Angeles, Chicago, New York, Atlanta and more, provide flights to almost anywhere in the world, including cities in the United States. Examples are shown in figures 12 and 13. For some unknown reason, airports within almost all of the cities in South-America fly solely to Miami. An example is shown in figure 14.

In nearly all cases, it is possible to arrive at any airport within the USA while departing from any airport (that is present in our data) in the world, transferring only one time at most. For example, all flights from Moscow towards the United States (shown in figure 10), arrive in large cities and these large cities subsequently fly to almost any airport within in the United States. An example of a case where more than a single transfer is required, is for example a trip from Kuala Lumpur to the Gulfport in the United States. Kuala Lumpur does not fly to Atlanta, which covers Gulfport, shown in figure 15.

An example where multiple options for a one-transfer line do exist, is from Montreal towards Huntsville, shown in figure 16. Since we have multiple options to fly to Huntsville, we can choose which transfer airport we would like to use. The delays and airlines are shown in figure 17. Since from this figure Detroit clearly has the second lowest average delays, we pick Detroit and investigate Detroit's airline delays shown in figure 18. Airlines AS, B6 and OO seem to have significantly lower flight delays. As such we investigate AS and OO. It appears the airline AS (Alaska Airlines) has no flights from Detroit towards Huntsville, since no yellow bar appears, shown in figure 19. However, low flight delays airline OO (Skywest Airlines) do fly from Detroit to Huntsville and quite numerous as well. This can be seen from figure 20. It appears that flying from Motnreal towards Huntsville should be done by taking Skywest Airlines via Detroit.

Typically, large cities tend to have higher delays than small cities. Although one of the largest airports Atlanta is an exception. For instance, flying from Calgery in Canada towards Kona on Hawaii shows extremely low delays for Hawaiian transfer airports and high delays for Metro-poles like Los Angeles. Results are shown in figure 21.

Figure 10: Flights from Moscow to the United States.

Figure 11: Flights from cities in Mexico to the United States.

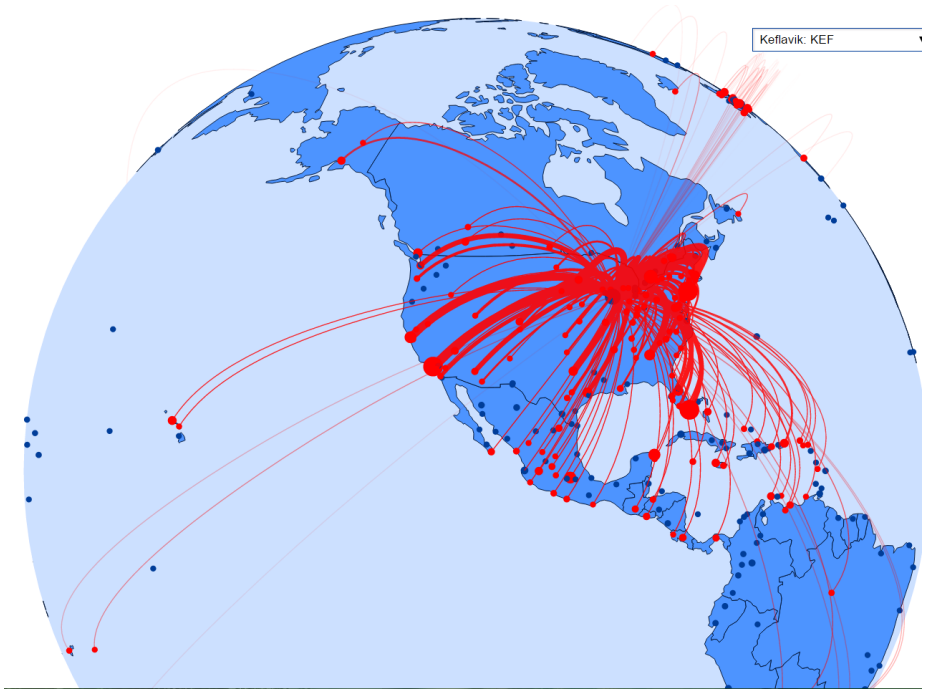

Figure 12: Flights from Chicago towards to rest of the world.

Figure 13: Flights from Atlanta towards the rest of the world.

Figure 14: Flights from South-America fly only to Miami.

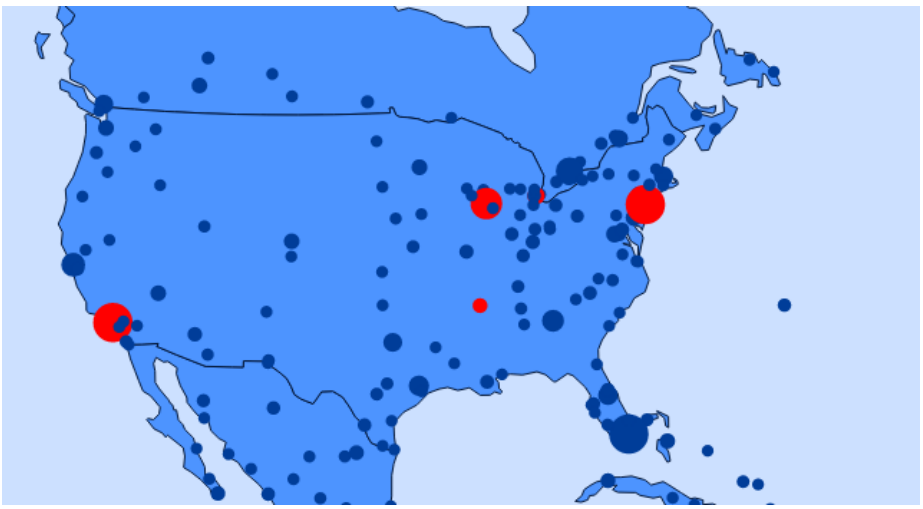

Figure 15: Departure airport: Kuala Lumpur.

Figure 16: One-transfer flights from Montreal to Huntsville.

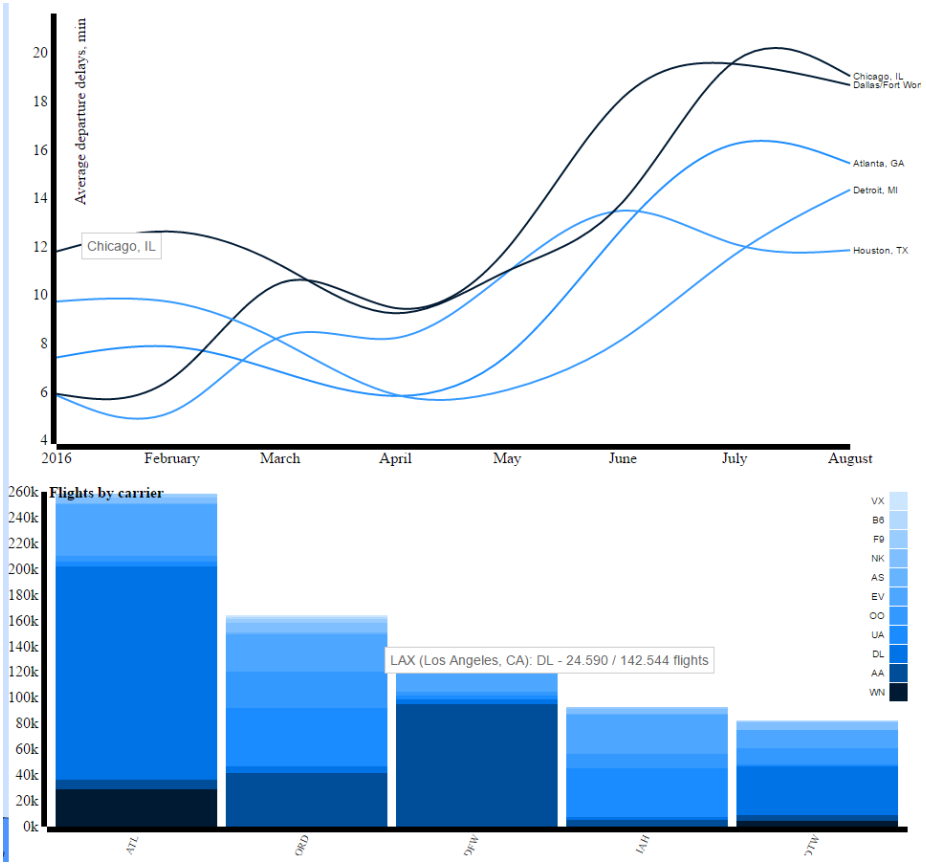

Figure 17: Flight delays and airlines for airports between Montreal and Huntsville.

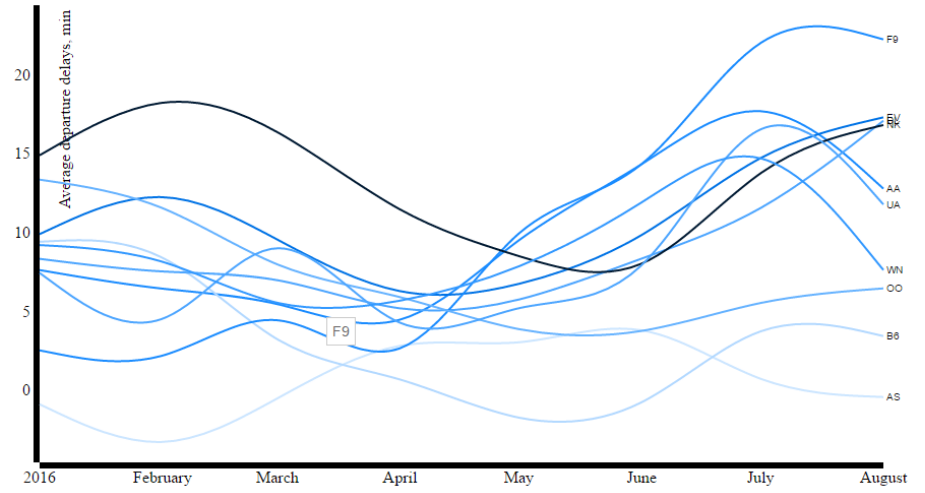

Figure 18: Airline delays for Detroit.

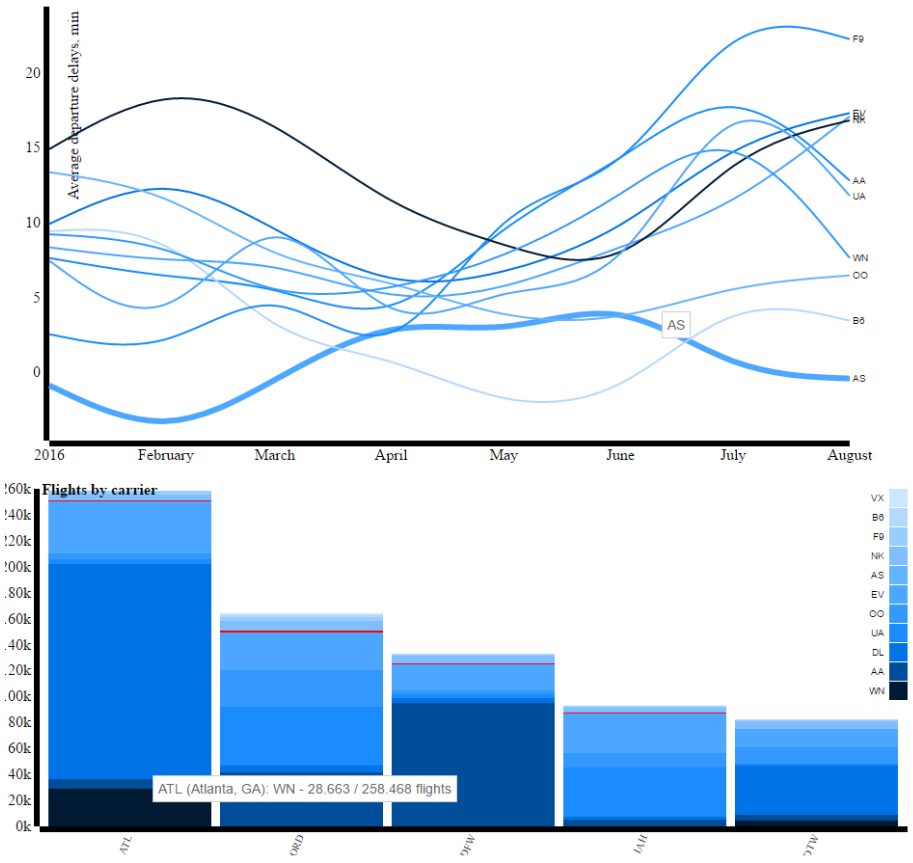

Figure 19: Flying from Montreal to Huntsville via Detroit, showing airline delays. Highlighted Airline AS (Alaska Airlines) delays.

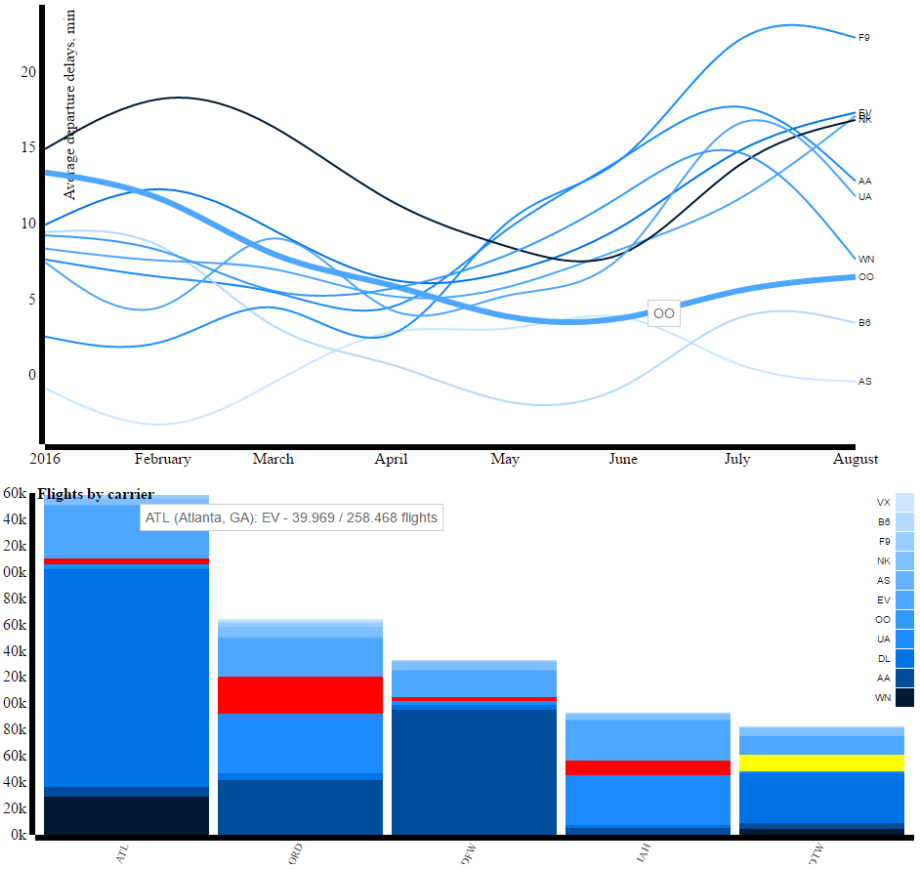

Figure 20: Flying from Montreal to Huntsville via Detroit, showing airline delays. Highlighted Airline OO (Skywest Airlines).

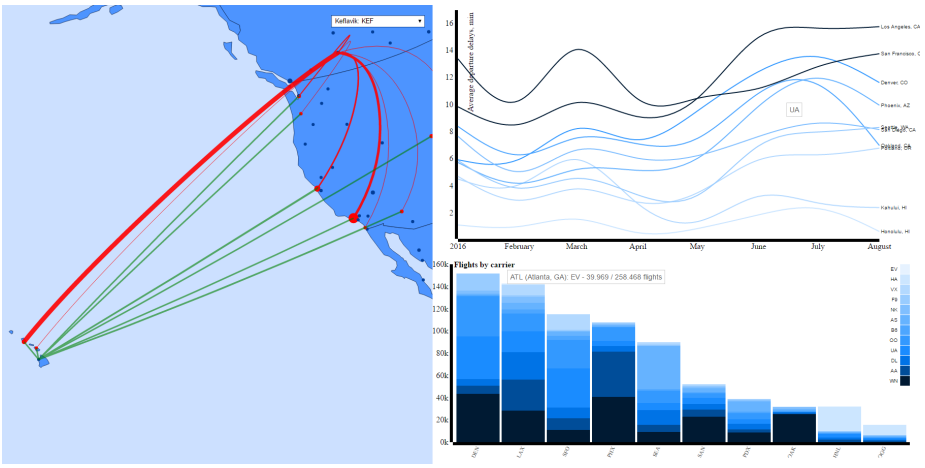

Figure 21: Flying from Calgery (Canada) to Kona (Hawaii). Large airports typically have significantly higher delays.

8 Suggestions for future Research

The application can be improved in many ways: Since the core user interface is present (the globe), all kinds of additional information regarding flights could be presented to the users. For example: The total ticket costs for each flight option could be displayed side-by-side so that the user can determine which flight option is the cheapest for him. We could not implement ourselves this feature because ticket costs were not present in our research data.

The application could also be extended by displaying the departure times of flights and flight durations to the user. Take the following scenario for example: The user wants to fly from X to Y. There is a direct flight with a flight duration of 2 hour and a ticket cost of $200. But, there could be a much cheaper indirect flight from X to Z to Y, which only costs $100, but the total flight duration is 6 hours. it will vary between travellers which flight option is preferable. Travellers could use the application to more easily discern their available options, or even filter flight durations on maximum flight duration and maximum ticket cost.

9 Appendix

9.1 Division of workload

The division of work within the group was fairly straightforward. It was decided that one programmed the visualization tool and reviewed the report, whilst the other gave his contribution on the techniques to use, the questions to be answered, recorded the screencast and later wrote the report.

References

[1] D3.js, `https://d3js.org/`, 2017.

[2] OpenFlights, *Airline Route Mapper Route Database*, `http://www.openflights.org/data.html`, January 2017.

[3] United States Department of Transportation, *On-Time : On-Time Performance*, `http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time`, November 2016.

[4] United States Department of Transportation, *On-Time : T-100 Market (All Carriers)*, `http://www.transtats.bts.gov/Fields.asp?Table_ID=292`, July 2016.