Dipartimento di Informatica, Sistemistica e Comunicazione
Master Degree - Data Science

Flights history

Candidates:
Luca Cilloni
Marta Privitera
Alice Parodi

ACADEMIC YEAR 2022/2023

Contents

Chapter 1

Introduction

The origins of air transport date back to the 18th century with the development of the hot air balloon, an apparatus capable of atmospheric displacement for entrainment purposes and few people transport. Then, a large step in progress came with the construction of the first powered airplane by the Wright brothers in the early 1900s. Since that time, aviation has been technologically revolutionized by the introduction of sophisticated jets, planes, rockets, helicopters, and drones. This fact led to undoubted economic implications: countries could actively participate to the global market by increasing participation to main markets, overcoming geographical barriers.

Indeed, through years aviation became a mean of international trade and development. In 2019, the air transport industry supported a total of 65.5 million jobs globally. Airlines, air navigation service providers and airports directly employ around three and a half million people. The civil aerospace sector (the manufacture of aircraft, systems and engines) employs 1.2 million people. A further 5.6 million people work in other airport positions.

One of the industries that relies most heavily on aviation is tourism. By facilitating tourism, air transport helps generate economic growth and alleviate poverty. Approximately 1.4 billion tourists are crossing borders every year, over half of whom travels to their destinations by air. The air technology has greatly increased in recent decades becoming one of the first mean of movement from one place to another. People have the possibility to access advanced, more comfortable form of transport and speed connections between cities. For this reason, it is natural to think that the number of passengers is growing over the years. To support that claim, we find the analysis on the evolution of passenger numbers in world history made by The International Civil Aviation Organization (ICAO). This specialized agency of civil aviation of the United Nations provides some useful information: the curve of passenger numbers turns out almost exponentially, undergoing an unprecedented decline just in 2020 with the advent of the pandemic COVID-19.

Chapter 2

Impacts of COVID-19

2.1 COVID-19 across air industries

Considering the rapidly spreading disease named as COVID-19, the International Civil Aviation Organization (ICAO) actively monitors its economic impacts on civil aviation and regularly publishes updated reports and adjusted forecasts. All analysis show that a decimation in demand among travelers has had a significant impact on the airline industry during the past few years. Significant reductions in passenger numbers have resulted in flights being cancelled or planes flying empty between airports, which in turn massively reduced revenues for airlines and forced many airlines to lay off employees or declare bankruptcy. The image 2.1, published by ICAO, shows the Global economic impact of Covid 19 in brief (from 2020 to 2021) highlighting the losses in those areas that the pandemic has affected most.

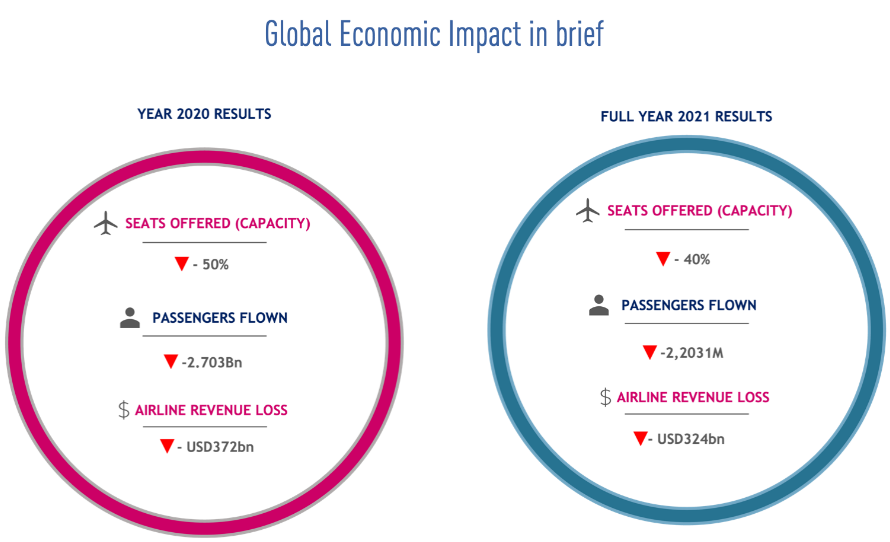


Figure 2.1

Government regulations in Europe and the United States mandated that airlines refunded fares when flights were cancelled or, in many cases, airlines offered vouchers or travel credits to avoid economic losses. Moreover, as passenger flights were cancelled, the cost of sending cargo by air increased rapidly and therefore also the international markets have suffered some important losses. On the contrary, business aviation field had an opportunity to capture premium passengers, who might previously have chosen airlines, that now may prefer the social distancing afforded by a private jet.

It is clear from the previous paragraphs that the pandemic halted activity at airports around the world. Airports Council International (ACI) World has published an assessment analyzing the impact of the COVID-19 pandemic on air traffic and below are described briefly some of the crucial points of it. Moreover, at the beginning of pandemic, ACI developed a projected model for air traffic changes in the different countries all over the world.

2.2 Global Overview on 2020-2021 from ACI

- Over the first two years of the pandemic, COVID-19 outbreak reduced the number of passengers at the world's airports by 11.3 billion.
- Global passenger traffic in 2021 has reached less than half of what it was in 2019 (only 4.4 billion of the 9.2 billion passengers served in 2018-19).

2.2.1 Flights in different countries during the pandemic

- The conditions of the United States and Canada, the two large and dominant economies of America, were not great. Both have suffered significantly due to the economic loss from the dissemination of the virus to every corner of their territories. Fueled by a fast-recovering US domestic market and a high overall vaccination rate, North America's performance has improved significantly in 2021. As a result, these regions better perfomed compared to all the others with their traffic down by only 37.3% compared to the projected baseline presented by ACI.
- Most of the smaller central American and Caribbean nations have suffered catastrophic economic hardships during the pandemic. The reason is that these smaller countries have limited natural resources, bad economy, poor governance and high levels of corruption. The Latin America-Caribbean region had a strong second half of 2021, bringing the region to report a decline of 43.1% compared to the projected baseline.
- Most countries in Africa confirmed their first COVID-19 cases within a similar timeframe as in Europe, and while the first cases were expected to arrive from Asia, these reportedly arrived from Europe and North America, and quickly spread throughout the continent, triggering an unprecedented halt in both international and domestic aviation and, with it, most tourism and travel-related activities. The intercontinental connectivity of some African countries with other regions as well as the intra-continental connectivity patterns, served as drivers for the spread. Africa recorded some significant gains just in the last quarter of 2021, which helped the region to end 2021 with a decline of 56.3% compared to the projected baseline from the previous year.

- The global COVID-19 pandemic arrived in Europe with its first confirmed case in Bordeaux, France, on 24 January 2020, and subsequently spread widely across the continent. While Europe recorded significant improvement in the third and fourth quarters of 2021, the region ended the year 60.1% below the projected baseline. Most affected countries in Europe were France, Germany, United Kingdom and Russia.

- COVID-19 arrived early to the Middle East region, with Iran recording the first cases in February 2020. These quickly spread to neighbouring states of the Gulf through business contacts and religious tourism. The effects of Covid-19 have varied widely across the region. Some states, including Israel and the Gulf states dealt relatively well with early containment and prevention measures followed up by extensive vaccination programmes. The Middle East remained the most impacted region in 2021, losing 65.9% of its passenger traffic for the year compared to the projected baseline. This is due to the region's high dependence on international travelling and connectivity, both of which are recovering much more slowly than domestic ones.

- After the virus was first detected in Wuhan, China, the number of COVID-19 cases across the Asia-Pacific region increased dramatically. Following an early start of recovery, Asia-Pacific experienced a slower than expected first quarter in 2021 as well as a sharp decline in domestic traffic in the third quarter, after travel restrictions were reinstated in some countries. The region ended the year with an estimated traffic loss of 62.7% compared to the projected baseline.

2.2.2 Domestic and international flights

- International passenger traffic volume lagged significantly behind domestic traffic recovery due to border closures by countries to stem the pandemic. In 2021, international travelers number was 27% of the 2019 level. The Omicron wave at the end of 2021 also contributed to slow the recovery of traffic.

- Domestic passenger traffic recovered faster than international traffic. Domestic travelers number were 61% of the 2019 level. Major domestic markets began recovering in 2020 and the pace of recovery accelerated in 2021 especially in the US, the world's largest domestic market. On the other hand, large domestic markets in Asia-Pacific have seen some setbacks resulting from the restablishment of travel restrictions. While the US and Russian domestic markets have recovered, the same is not true for the other major domestic markets of China, Canada, Japan and Australia.

In the last year, the speed of the recovery continued to depend on the individual country involved and on the level of coordination pursued by national governments worldwide. While the global international travel market was still mostly depressed, a growing number of countries was moving towards the gradual reopening of their borders to vaccinated travelers.

What is clear is that the recovery is uneven. Domestic markets are recovering at a much greater speed than international passenger markets. However, this depends on the specific region or country of the world. For instance, the rapid recovery of the Chinese domestic air passenger market, was severely dampened in the second half of 2021 by a resurgence of COVID-19 and the re-establishment of travel restrictions in the region.

Chapter 3

Specific Questions and Data Set Description

3.1 Specific Questions

- How has air traffic industry been developed across the last decades?
- What and how has changed with COVID-19 impact?
- Were this changes different across the continents?
- Were there other important trends to be analysed?

3.2 Data Set Description

In order to answer questions of interest, we worked with four different data sets that are connected to each other because they are point of view of the same story.

3.2.1 Airlines Traffic Passenger Statistics

This data set has been found in Kaggle. It contains information on air traffic passenger statistics from 2006 to 2015. There are 17 columns:

- Activity Period, the date of the activity Operating Airline, the airline that operated the flight
- Operating Airline IATA Code, the code of the airline that operated the flight
- Published Airline, the airline that published the fare for the flight
- Published Airline IATA Code, the IATA code of the airline that published the fare for the flight
- GEO Summary, a summary of the geographic region (domestic or international)
- GEO Region, the geographic region
- Activity Type Code, the type of activity (deplaned, enplaned or thru/transit)

- Price Category Code, the price category of the fare (low fare or other)
- Terminal, the terminal of the flight (Terminal 1, Terminal 2, Terminal 3, International or Other)
- Boarding Area, the boarding area of the flight (A,B,C,D,E,F,G or Other)
- Passenger Count, the number of passengers on the flight
- Adjusted Activity Type Code, the type of activity adjusted for missing data
- Adjusted Passenger Count, the number of passengers on the flight adjusted for missing data
- Year, the year of the activity
- Month, the month of the activity

Analyzing these variables, we noticed few things to be fixed. First, *Activity period* column is redundant because it is the combination of the year and the month, such as 200507, but the columns related to the date are already present. Columns *Operating Airline* and *Published Airline* usually coincide, but for few observations they are different, so we decided to keep them. On the other hand, we delete both *passenger count* and *activity type code* because we already have the adjusted ones after having checked the missing values. Moreover, the codes and the names of the airline are two columns that describe the same feature, so there is not a loss of information if we delete one of these. At the end of the cleaning process, data includes the number of passengers, the operating airline, the published airline, the geographic region, the geographic summary, the activity type code, the price category code, the terminal, the boarding area, and the year and month of the flight.

3.2.2 Air transport of passengers

This second data set from Eurostat has been added to construct a connection between 2015 and COVID period. Thank to that, we have the whole vision of the changes in air traffic during the last decades. Few different variables are stored, but just three of them will be useful for our study:

- geo, the code of the geographic region
- time_period, the year of which the data are related to
- obs_value, the number of passengers in that year and geographic region

3.2.3 COVID-19 Air Traffic Data

This third data set is a sort of picture of how the development of air traffic changed during the COVID-19 pandemic. It is derived from the full OpenSky data set, where all the information about flights is stored. It includes all flights around the world between January 2020 and May 2020, to emphasize the sudden variation in those months. It contains the following variables:

- Country, the name of each country considered
- Day, the date of each flight
- Domestic_flights, the number of flights that have the departure and the arrival in that specific countries in that specific day
- International_departures, the number of departures of the flights that leave that country in that day
- International_arrivals, the number of departures of the flights that arrive to that country in that day
- Total_flights, the sum of the three variables before

There are 11 154 observations and 6 variables. It will be easily seen the path according to the geographical area of the flights during this transition phase.

3.2.4 Air traffic for each airport in the world

To make our project complete, we decided to zoom in to particular cities around the world to see how the number of flights changed between 2019 and 2020. Also here, data are extracted from the full OpenSky dataset, where data have been periodically included in the data set. The origin and destination airports are computed online based on the ADS-B trajectories, so no external sources of data have been taken into account. Since each file is related to a specific month, the files have been merged according to the origin airport. To summarize, we list all the attributes:

- Callsign, the identifier of the flight displayed on the screens
- Number, the commercial number of the flight
- Icao24, the transponder unique identification number
- Registration, the aircraft tail number
- Typecode, the aircraft model type
- Origin, a four-letter code for the origin airport of the flight
- Destination, a four-letter code for the destination airport of the flight
- Firstseen, the UTC timestamp of the first message received by the OpenSky Network
- Lastseen, the UTC timestamp of the last message received by the OpenSky Network
- Day, the UTC Day of the last message received by the OpenSky Network

These data are very heavy, consequently we made a sample of the population of flights starting from a specific city. In particular, we made a systematic sampling: a probability sampling method where researchers select members of the population at a regular interval. In our example, this is translated into the fact of sampling the first 2 days for each month. This can be done because we already have a natural order of the units given by the time and the goal is to be as efficient as possible. This approach tries to maximize distance between units in order to have a high variance within the sample so that the variance of the estimator is smaller.

Chapter 4

Graphic Design

Based on data sets just described, the infographic we create is realized with several charts arranged on several pages. In this chapter we analyze in detail the plots that make our data visualization up and what we aim to communicate through them.

4.1 The goal of the analysis

The infographic aims to represent the change in the number of aircraft passengers and the number of flights over the years. We want to understand how these numbers are changing and how the pandemic has affected their evolution over time. We also intend to go into detail and represent the differences in flights between states during the pre and post COVID period. In particular we will focus our attention on a specific city to support our thesis.

4.2 Target

The infographic is aimed at all those interested in the phenomenon, without demanding any kind of previous knowledge. In relation to the target to which it is addressed, the infographic will be created in such a way a simple, comprehensible reading is allowed so that it can be accessible by any user. Our analysis can be a useful tool for understanding the airline industry of the last past years. From an expert point of view, the data visualization could be used to study the future effects of different factors on air traffic passengers number, such as the time of year or day and the price of airfare.

4.3 Estimated vs Actual air passengers trend

The first visualisation of our infographic illustrates the trend of the yearly number of flying passengers over a twelve-year period, starting in the year 2010, which is the beginning of our time horizon, and ending in 2021. In the chart 4.1 we can appreciate the estimated growth of the trend in the years 2020 and 2021 in accordance with the forecast and the relative margin of uncertainty, up to a peak value of more than 5 billion 450 million annual passengers reached in 2021. In the graph below, on the other hand, we witness the actual situation, where due to Covid a minimum of about 670 million was reached in 2020, accompanied by a slight upturn over the following year. The intention was to highlight the covid-related drop in the number of total travellers.

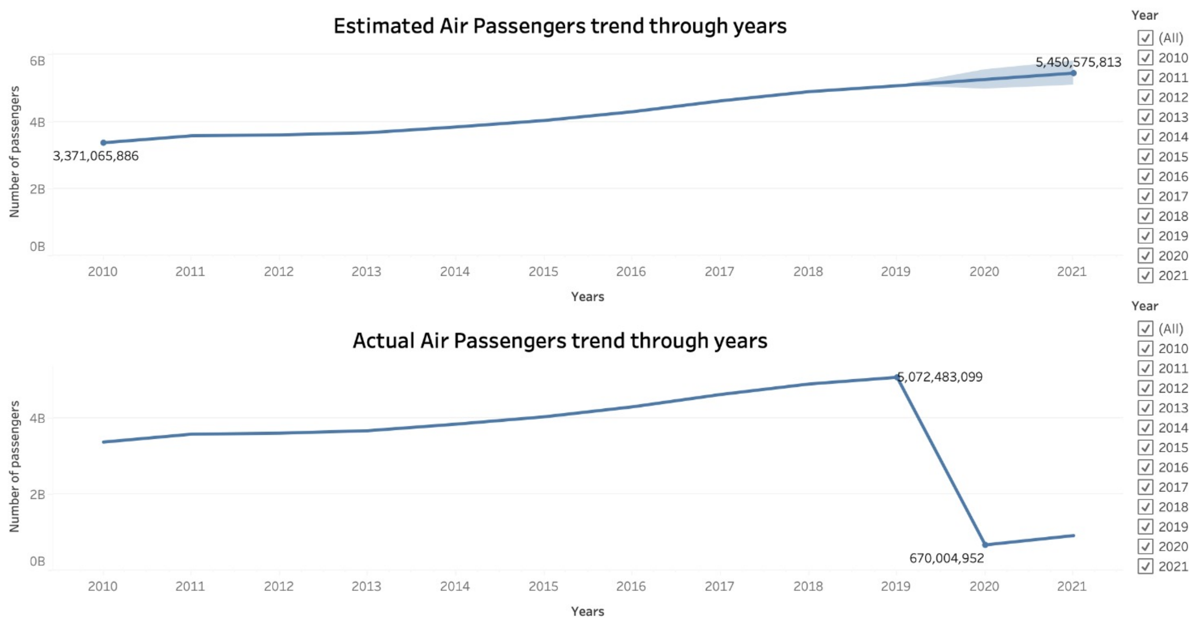


Figure 4.1

We visualize data by choosing line as visual encoding. The contrast between the red of the title and the blue of the line makes all the information we want to communicate readable.

We divide the visualization into two separate parts. The top part represents a pre-COVID estimate: the user can change the filter and see the difference in the estimate from 2020 to 2021 based on the selected years 4.2. Similarly, a filter in the right column is provided for the second chart (post COVID result) in order to see how the maximum and minimum number of passengers vary if it is decided not to consider 2019 or 2021. The filter can also be useful to zoom the trend over the range of years of interest, while maintaining the proportions initially established according to the rule of thumb.

Figure 4.2: Dynamic filter of years

4.4 COVID-19 evolution

The second infographic (see 4.3) is a dashboard, it is constituted by two bubble graphs. In each one there is a redundancy as the inner percentage, the value and the size of the bubbles indicate the same magnitude, in order to facilitate the user and render the reading immediate. On the left side, the evolution of the total number of flights from January to April 2020 is shown, whereas the right side depicts the difference in percentage compared to the previous month. It is interesting to point out that while in Europe, Africa, and America, the number of flights in February remains more or less the same or has even increased, as they are not yet affected by COVID, we see that in Asia (and to a lesser extent in Oceania) there has already been a drop of 14.8%. Moving forward to the month of April, it is possible to observe the almost total drop in Africa (-99.9%) as a consequence of the closure of the borders and the fact that most of the flights operated on the African continent are international. It is followed by Oceania (-93%) and Europe (-92.3%), while as regards America, where domestic flights are much more used - we will see a more in-depth analysis on the USA in the next viewings - the fall stops at 67.2%.

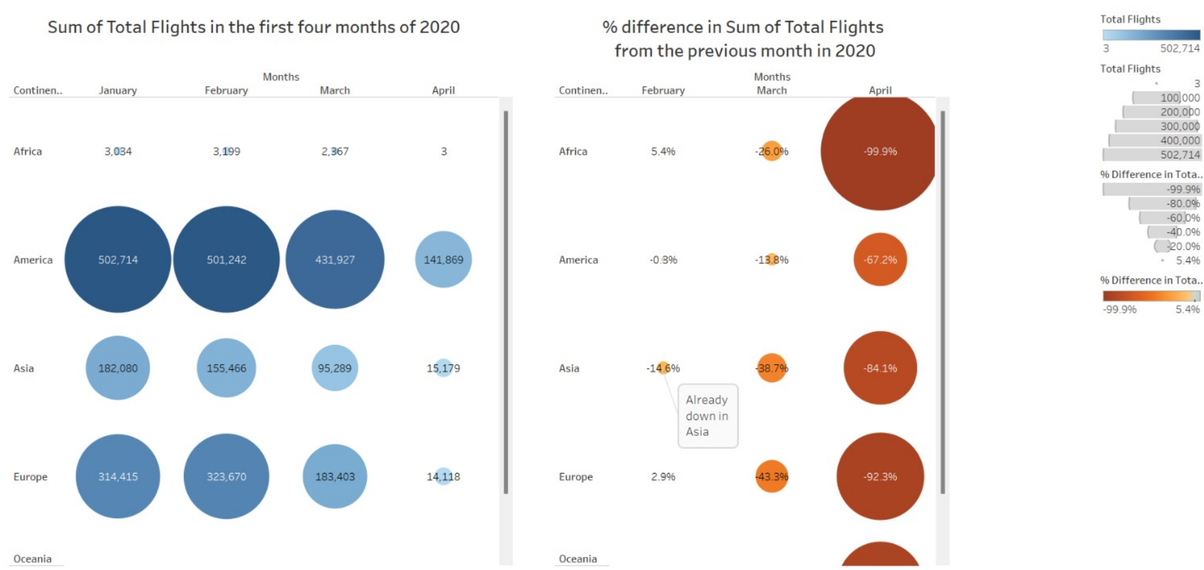


Figure 4.3

In this case the filters are not dynamic. There is a legend in which the size of the bubbles and the range of variation of value are shown (fig.4.4). Contextual elements like labels are also added on the bubbles to further emphasize the difference in quantities.

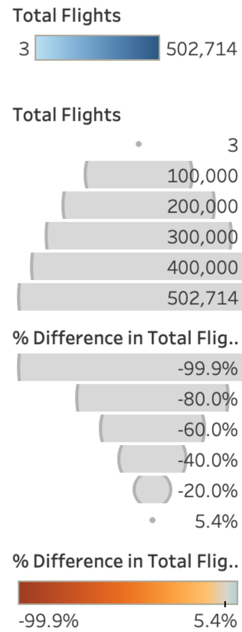


Figure 4.4: Legend for bubbles and value

4.5 Focus on flights from Rome 2019-2020

In the third rendering (figure 4.5) a more specific focus is placed on a single city, in this case we chose Rome. Thanks to the visualisation, it is possible to see graphically how the number of flights has changed from 2019 to 2020. The data was taken from the OpenSky data set, from which a systematic statistical sampling was subsequently done to reduce the workload handled by Tableau, using the first and second day of each month of the two respective years under examination. Each line represents a single route departing from Rome, and, at first glance, it is visible the difference in the number of departing aircraft, which has been almost halved from 17 452 to 10 391. Furthermore, by highlighting an individual route with the cursor, it becomes possible to see the number of trips made on the same route in the following year.

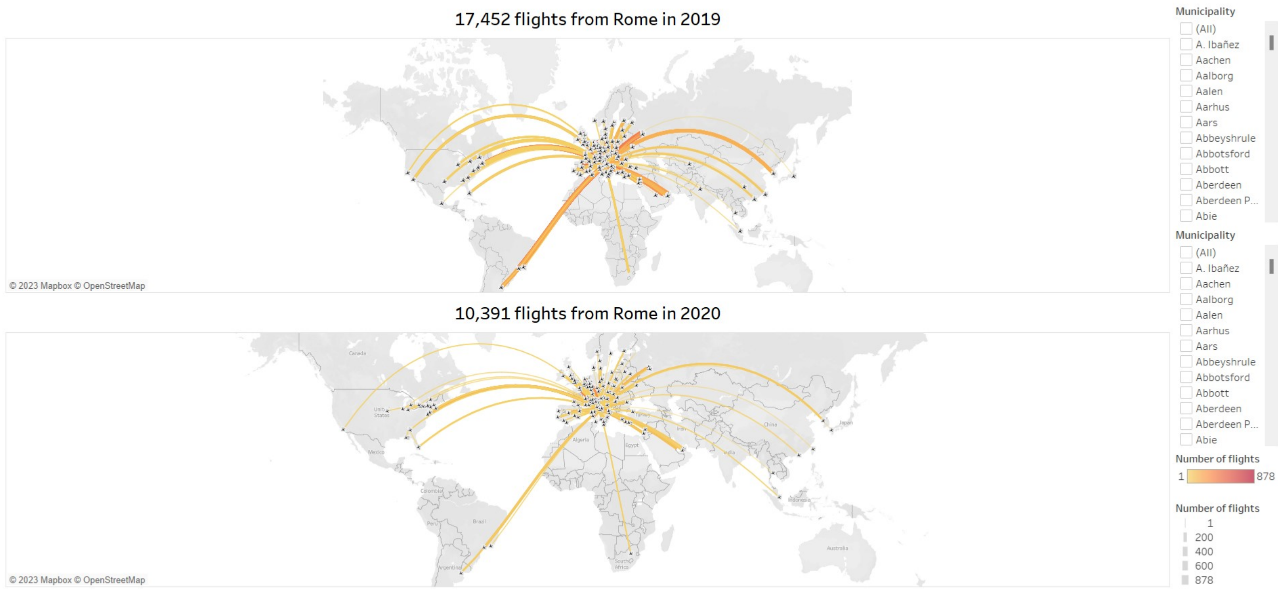
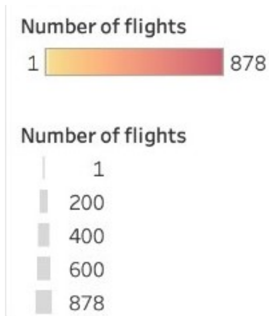
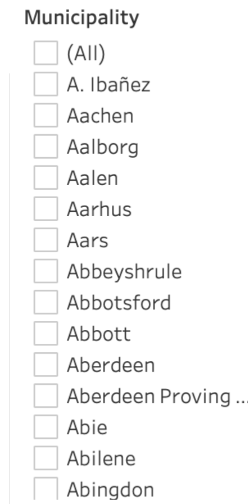


Figure 4.5

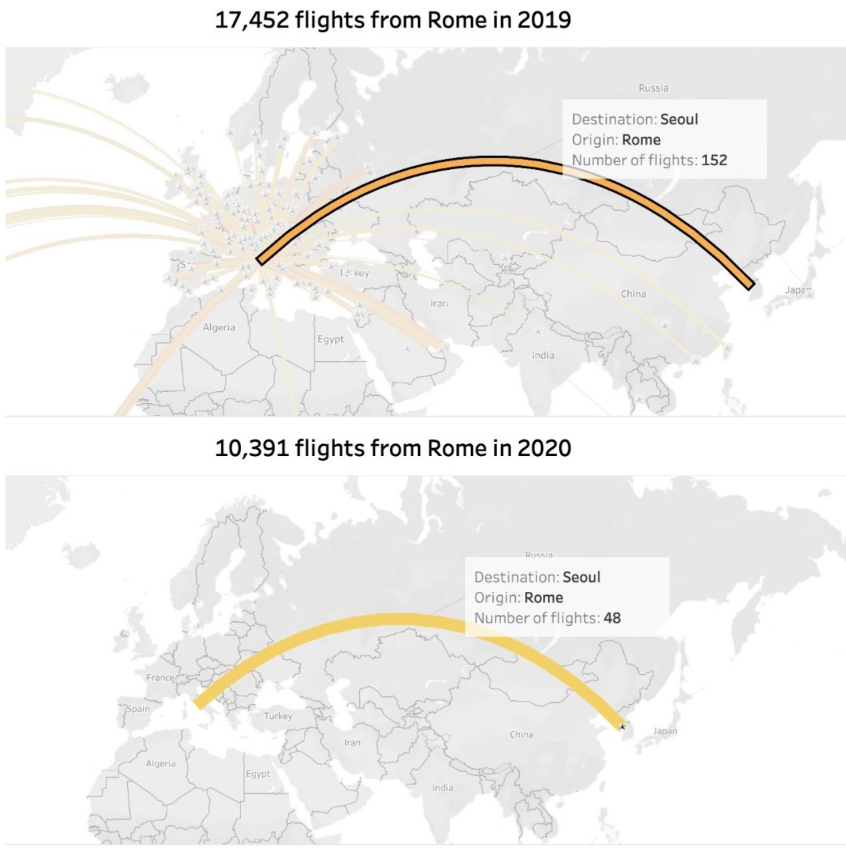
Here too, different visual encoding are used: the thickness of the lines indicates the density of flights departing from a specific city, the value as concentration of flights in different orange shades. It is possible to filter by city (dynamic selection in 4.6b). Moreover, there is a legend by color and size of lines (in 4.6a). The user can also select an aerial trajectory in the top or bottom chart and see in correspondence of it the change in terms of thickness and color of the corresponding trajectory in the other chart (see4.6c).



(a)



(b)



(c)

4.6 Seasonality of air passengers

At this point, we addressed another trend that was revealed by our analysis on the 4.7 plot: a distinct seasonality, which results in the number of flights being lower in February, before peaking in the summer months of July and August. A further observation is that the seasonality is mostly present in Europe and among the northern hemisphere, while in continents located in the southern hemisphere such as Australia and South America, where the seasons are reversed, this does not occur. Moreover, from 2005 to 2016, it is apparent that there is also an annual trend where flights are slightly increasing. Using this visualisation, we are able to simultaneously analyse both the monthly trend and annual growth in the number of flights. The latter is depicted through dots in different colours, where each dot corresponds to a year, while the months are displayed on the x-axis, with each box plot corresponding to a month of the year.

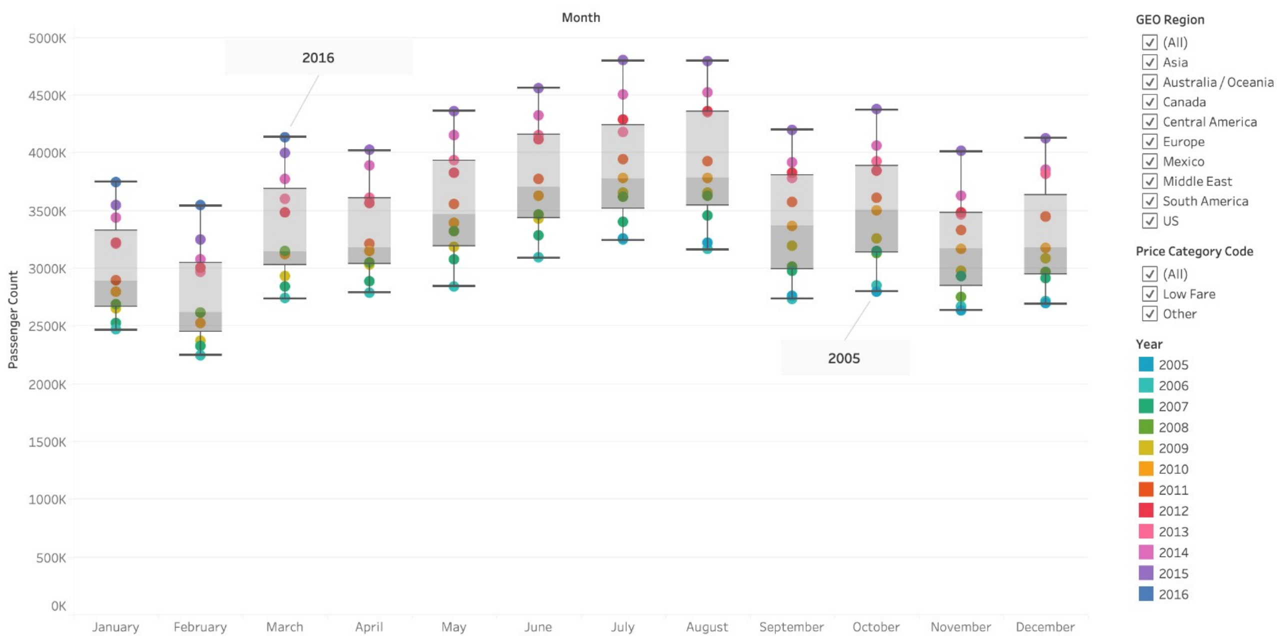
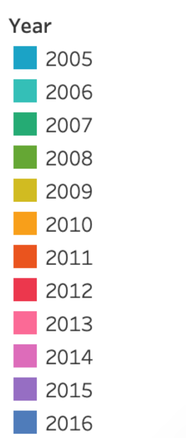
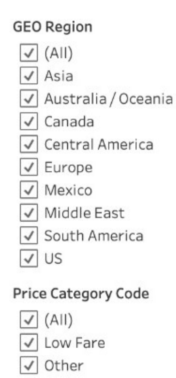


Figure 4.7

On the right side a legend is placed to show the different colors associated to the points in the graph. We choose a rainbow palette to best distinguish one specific year from all the others 4.8a. A dynamic filter allows to select all regions or a specific one (4.8b). In addition the user can choose the desired price category and see how the number of passengers changed over time depending on this factor.



(a)



(b)

4.7 International vs Domestic flights

Through the graph 4.9 we are comparing data on international flights (top part), computed as the sum of arrivals and departures, and numbers representing domestic flights (bottom part). The situation concerns the first five months of 2020, where we can see a huge drop in China, where the sum of international and domestic flights barely reaches 2,500, while the airports of Canton, Chengdu, and Shenzhen are usually among the most trafficked in the world. Subsequently, it is remarkable that in Europe, the countries with the highest number of international movements are Germany and the United Kingdom, where in the latter people are flying abroad over 11 times more compared to the national flights. Although the EU is less than half the size of the US, it has almost 200 million more inhabitants. Despite this difference, the number of passengers on flights in the EU and the US is almost the same, but passenger habits and types of flights are quite different. Air traffic within the US is massive, mainly due to the large number of domestic short-haul flights, since, as we see, 79,43% of all domestic flights around the world occur in the US. Europe's smaller size and higher population density make the train or car a more suitable travel option in many cases, influencing people's habits and reducing air travel. Moreover, this could be explained by the absence of alternative means of transport, the American railway network is useful for the transport of goods but as far as passengers are concerned, the transport is minimal, due to the inadequacy of the infrastructures.

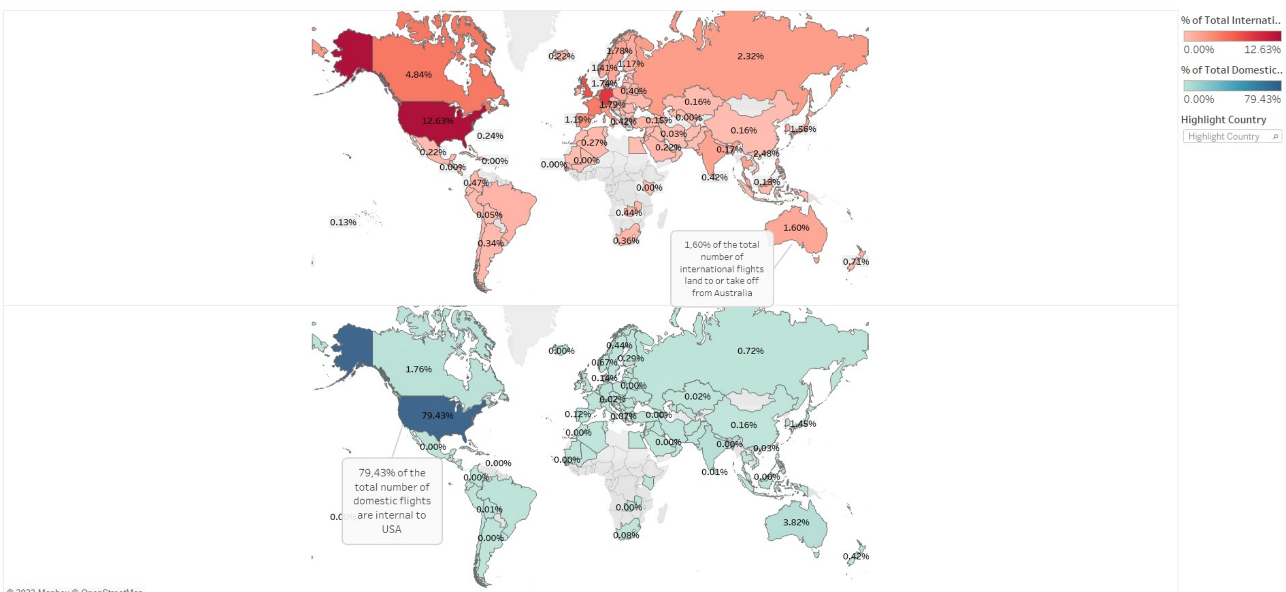


Figure 4.9

A legend on the right side highlights the use of colors in the maps: darker shades of colors mean greater quantities while lighter shades are associated to smaller quantities. Through the legend the distinction between the two categories of flights is clear: the warm colors concern the percentages of international flights while the coldest ones represent the percentages of domestic flights (see 4.10). In addition, a dynamic filter allows to select a country and see the differences in percentages between its international and domestic flights.

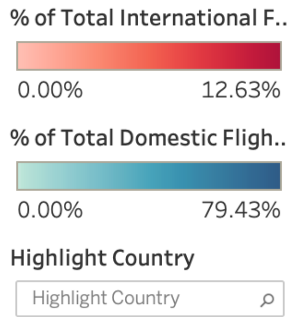


Figure 4.10: Legend fo international and domestic flights

4.8 Flights infographic history

In our analysis, the main objective is to assess the impact that the pandemic has had in various areas of the aviation industry, in particular on the number of passengers and aircraft. We also aim to underline the evolution in time of aviation industry. To achieve our research goals we used Tableau program and we set English as the writing language. This chapter describes and illustrates the infographic created using the plots presented in the previous chapter.

4.8.1 Overview first, zoom and filter, then details on demand

In agreement with what formulated in the mantra of data visualization, our story first deals with providing a general view on the total number of passengers over time. Then it goes into detail of individual countries by evaluating the percentage difference of flights from one month to the other in 2020. Finally we zoom in on a specific city to show the change in the number of flights from 2019 to 2020 with the arrival of COVID.

We also decide to offer an overview on the development of the aviation industry over time, focusing on the increase in the number of passengers in a period of time that goes from 2005 to 2016. We conclude the story by highlighting the huge difference between the number of flights (both domestic and international) in USA and those in the rest of the world.

Chapter 5

Heuristic Evaluation

The heuristic evaluation involved 4 expert users who were asked to freely interact for a few minutes with the data visualization, commenting aloud what they were doing and understanding. By interpreting user behavior and comments, we have identified some issues related to our data visualization.

5.1 Plot 1

Longitudinal Comparative Assessment

The comparison between the two versions of the same data visualization is shown in the figure 5.1. During the interaction it emerges that the first plot related to the estimated versus actual number of air passengers is clearer if the trend of the number of air passengers in the world is displayed and the filter on a specific area of departure flights is not specified. The main reason why the top plot is not clear is that the filter column shows just the country code. In this way it is not easy to spot the name of the country and complete the task, while in the general overview this problem doesn't exist at all.

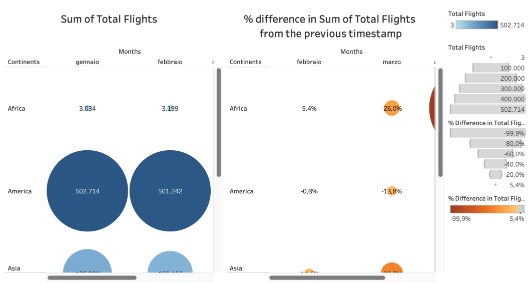
All the users agree with the fact that because COVID has affected all nations of the world in a similar way, discrimination by state is secondary to the view of the change in the trend. For this reason the plot without filter is judged better than the one with filter.

Figure 5.1: Comparison between the two versions of the same line plot. The first plot shows the number of passengers per country, the second one the number of passengers in the world.

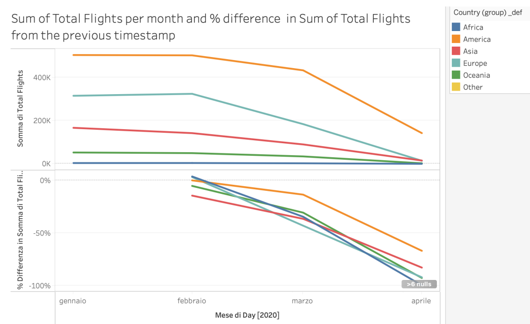
5.2 Plot 2

Cross-sectional Comparative Assessment

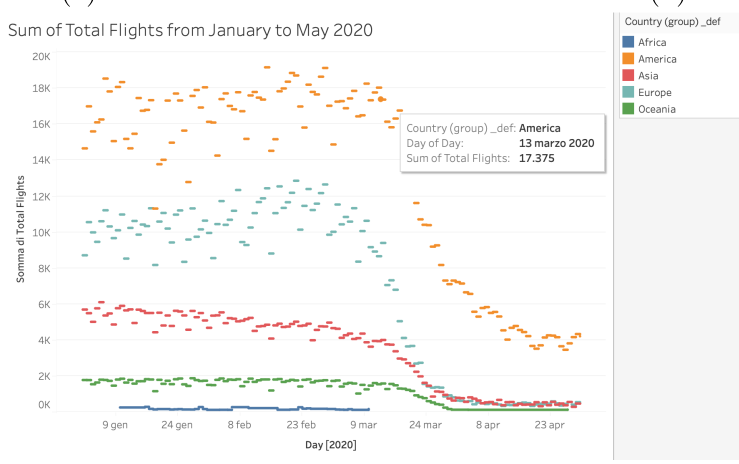
The second plot related to the trend of COVID-19 in different continents of the world (in the period of time that goes from January to April 2020) is presented in different versions. In the first version 5.2a there is redundancy made by label, size and value. This helps users to catch the information at a glance. The second chart in 5.2b is a line chart. To understand the percentage difference we are advised to invert the y axis and then make a growing and not decreasing line at a greater percentage difference. With the third plot (5.2c) it is possible to understand the trend as the days change, while the other graphs focus on months. For someone, it is also difficult to understand what the dashes mean. In general, the third version is considered more difficult to be interpreted than the others due to lack of contextual information, labels and tool tips. Since the first version is judged the best from every point of view, we decide to keep that version.



(a)



(b)



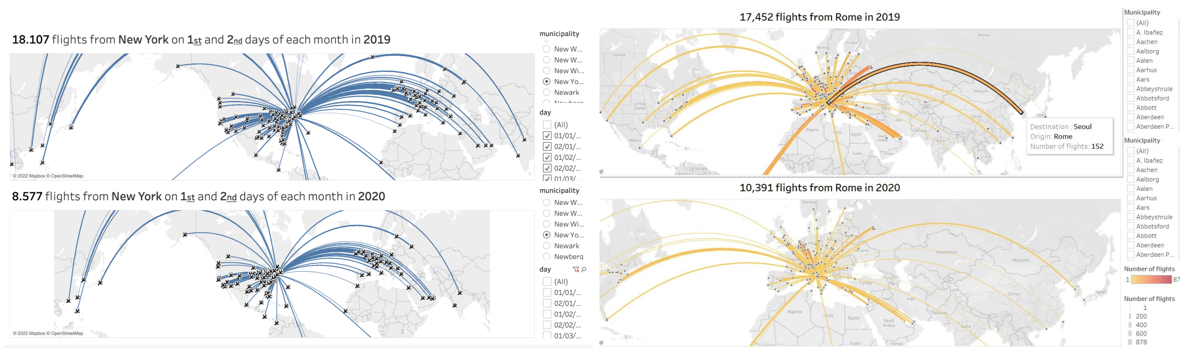
(c)

Figure 5.2: Top right: Percentage difference in line chart. Top left: Percentage difference in bubble chart. Bottom center: Percentage difference in dash chart.

5.3 Plot 3

Formative Assessment

The third graph shows a focus on a specific city: New York 5.3a. For all users it is not clear the difference between the two years. This is probably due to the fact that each flight trajectory is composed by numbers of superimposed lines. We are suggested to use value to show the difference in density in terms of color gradient. In addition, they required to add as a label on each trajectory line the number of flights for that specific trajectory. We decide then to follow suggestion and create the plot in 5.3b. To do that, we had to create a unique line for a specific path with different width and value according to the number of flights. Finally, we choose to focus on Rome instead of New York to emphasize more the phenomenon.



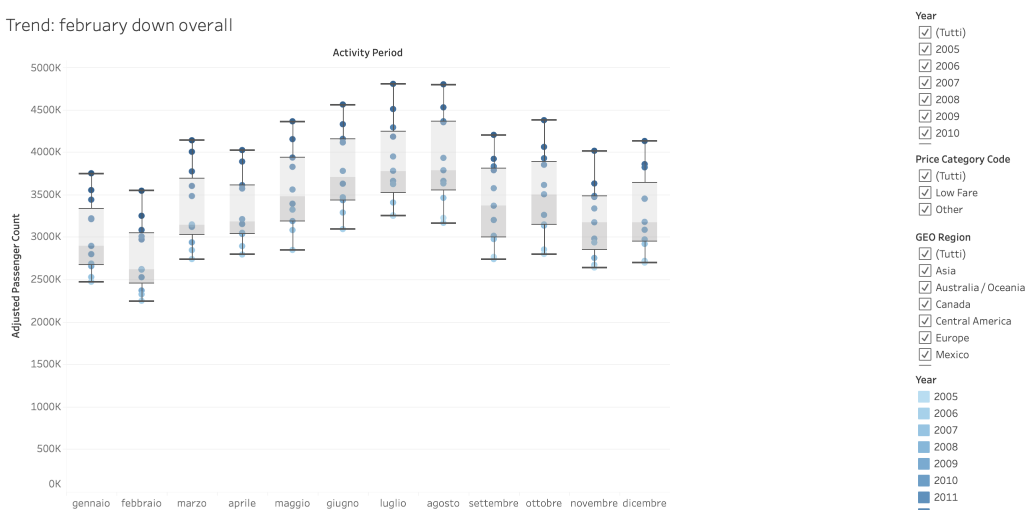
(a) (b)

Figure 5.3: Right image: one line is a flight. Left image: one line is density of flights for a specific path line.

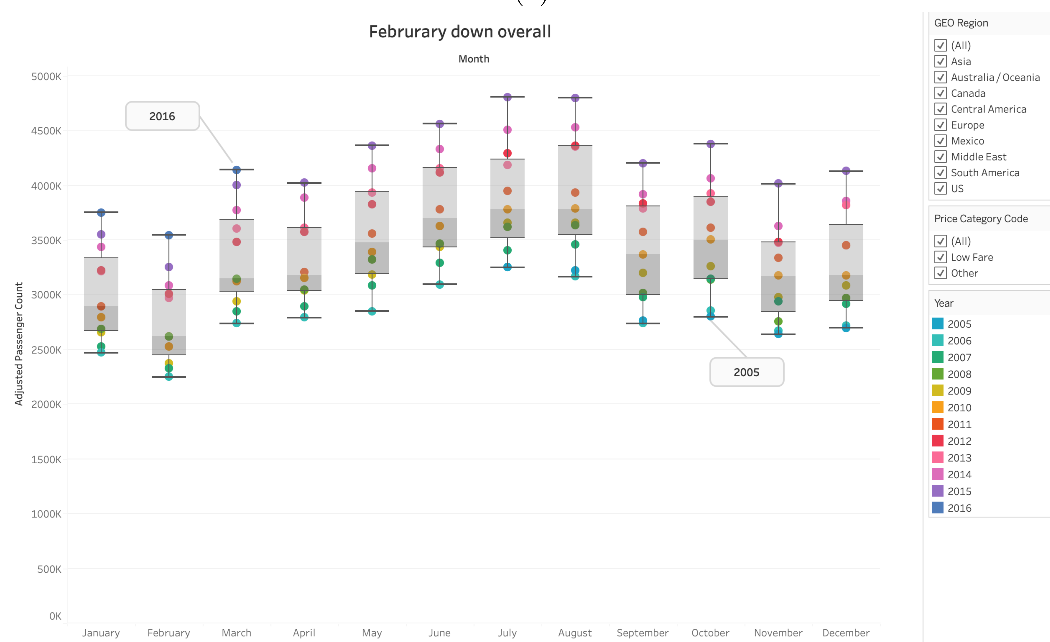
5.4 Plot 4

Formative Assessment

In the fourth plot 5.4a we show the increase in the number of passengers during the summer months with a decrease in the winter months (see February as an example). We are suggested to change the shades of color so that the transition from one year to another results clearer. We choose to divide the color range into multiple bins to have different color per year. Based on the users comments we also decide to add annotations to highlight the initial and final year and a filter to see the years individually (see 5.4b for changes).



(a)



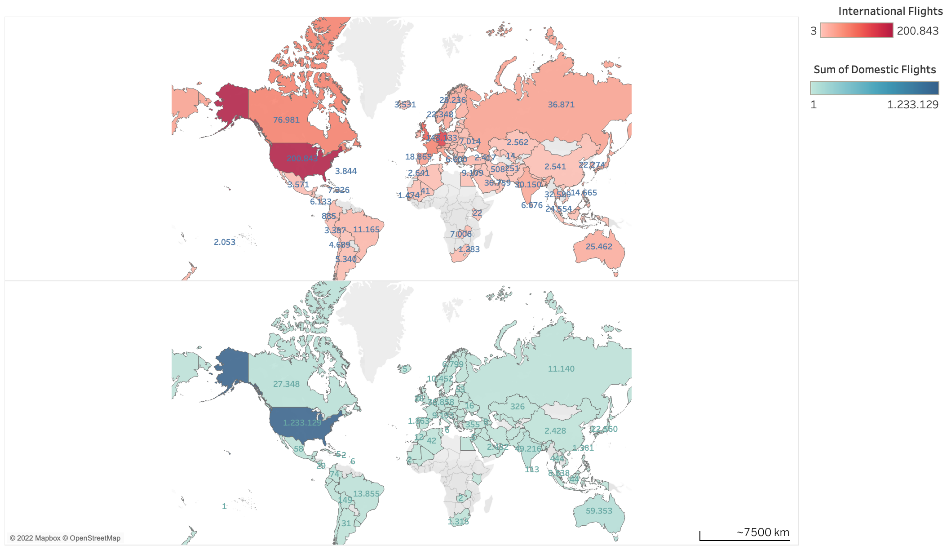
(b)

Figure 5.4: Left image: first version of boxplots. Left image: second version of boxplots.

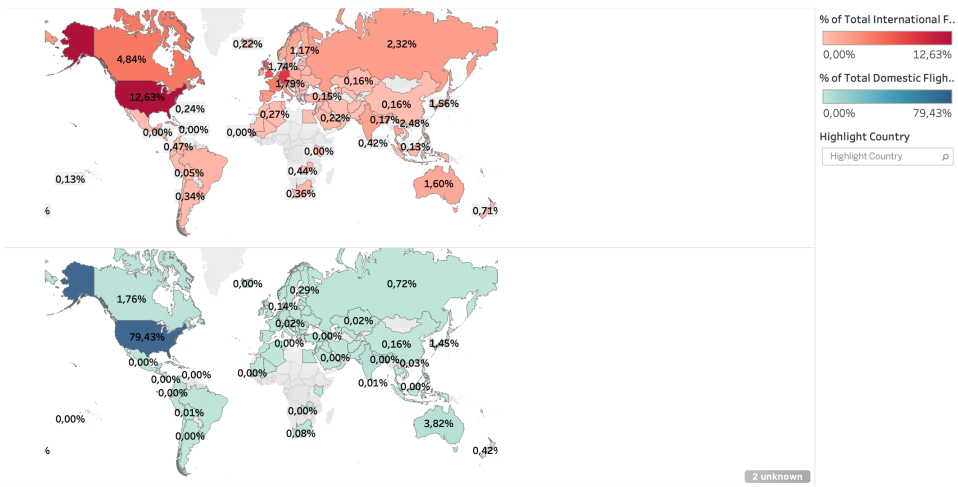
5.5 Plot 5

Formative Assessment

The plot shows choreopleth map of international and domestic flights 5.5a. Here we are advised to use percentages instead of absolute values (see 5.5b for changes).


International Flights
3
200.843
Sum of Domestic Flights
1
1.233.129
76.981
36.871
25.462
27.348
11.140
1.233.129
59.353
13.855
© 2022 Mapbox © OpenStreetMap
~7500 km

(a)


% of Total International F..
0,00%
12,63%
% of Total Domestic Fligh..
0,00%
79,43%
Highlight Country
Highlight Country
4,84%
12,63%
2,32%
1,60%
1,76%
79,43%
0,72%
3,82%
2 unknown

(b)

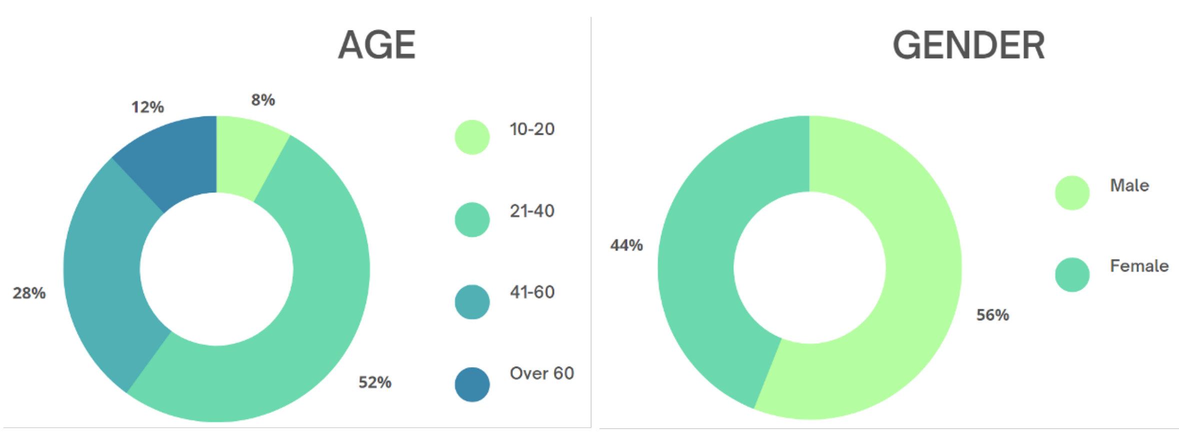
Chapter 6

Psychometric Questionnaire

After submitting a number of users to the heuristic evaluation and user test, it has been decided to administer a questionnaire to assess the quality of data visualization using the Cabitza-Locoro Scale. This assessment tool consists of two sections. In the first section, the user is asked to rate the quality of the infographic by means of a Likert scale ranging from 1 (very little) to 6 (very much) for each of the following qualities:

1. Usefulness
2. Clearness
3. Informativeness
4. Beauty
5. Intuitiveness

In the second section instead, the user is asked to rate, from 1 (very low) to 6 (very high), the overall perceived value of the data visualization. The target group to which the questionnaire is submitted is mostly concentrated in the 21-40 age bracket and equally distributed according to the gender:



(a) (b)

Figure 6.1

and consists of 60 % non-experienced users and 40 % experienced users:

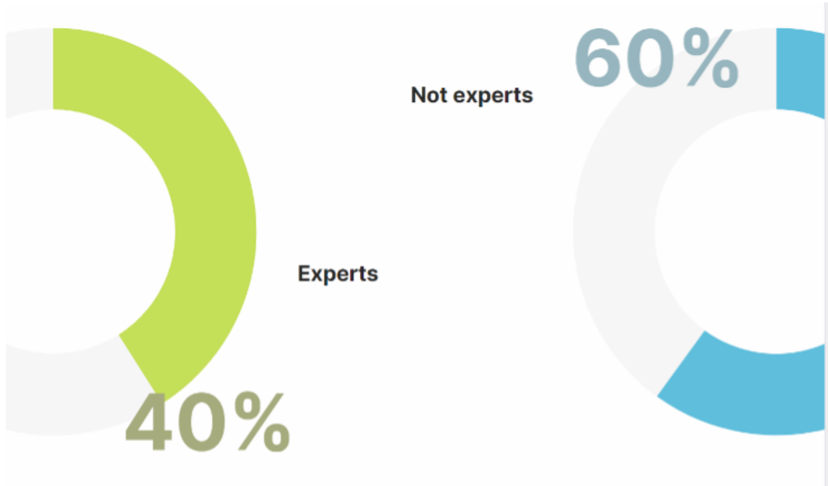


Figure 6.2

A total of 25 subjects participated in the questionnaire. Below is the analysis of the data collected:

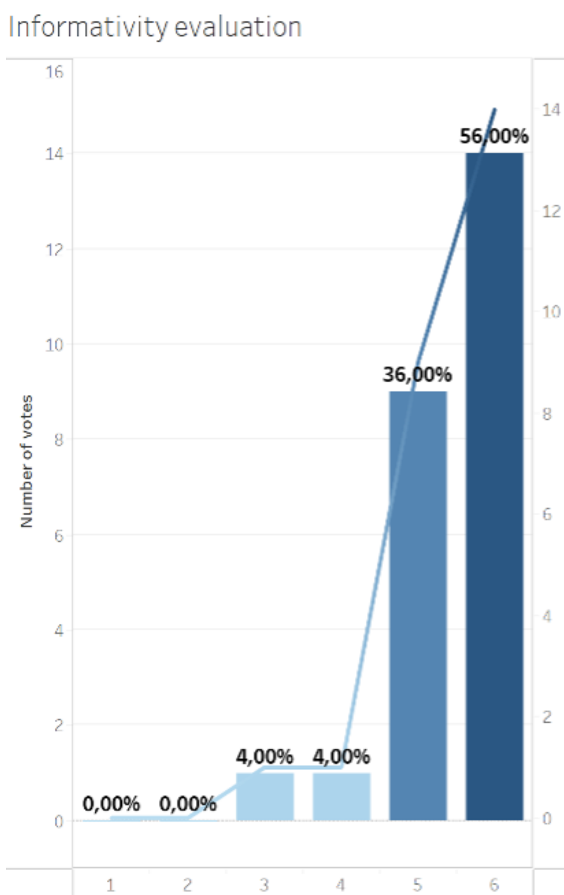


Figure 6.3

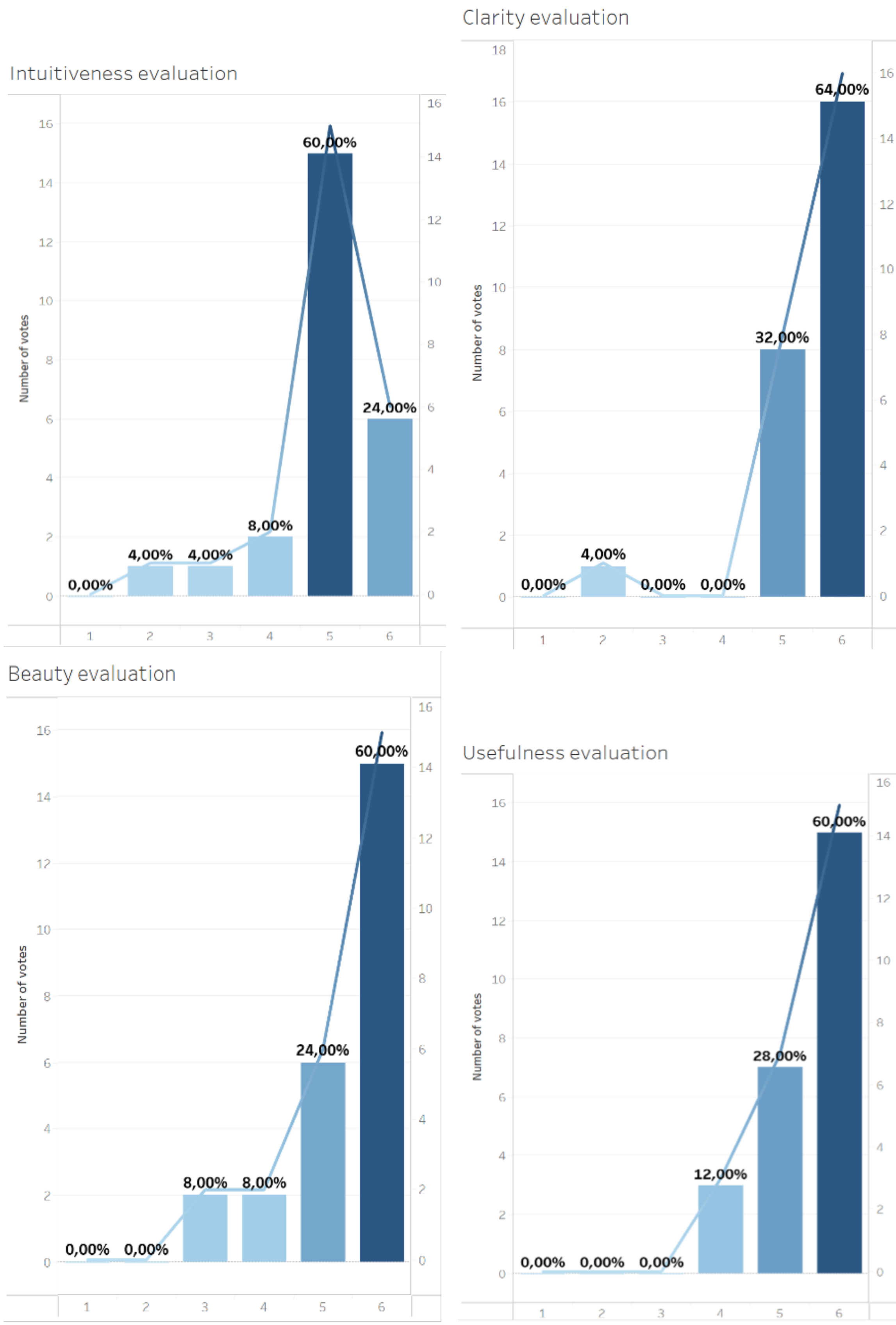

Intuitiveness evaluation
Number of votes
0,00%
4,00%
4,00%
8,00%
60,00%
24,00%
Clarity evaluation
Number of votes
0,00%
4,00%
0,00%
0,00%
32,00%
64,00%
Beauty evaluation
Number of votes
0,00%
0,00%
8,00%
8,00%
24,00%
60,00%
Usefulness evaluation
Number of votes
0,00%
0,00%
0,00%
12,00%
28,00%
60,00%

Figure 6.4

We can observe from the proposed results that the infographic was appreciated in all the evaluation fields. With regard to intuitiveness, it achieved a lower score than the other attributes, probably influenced by the fact that the user pool is predominantly inexperienced.

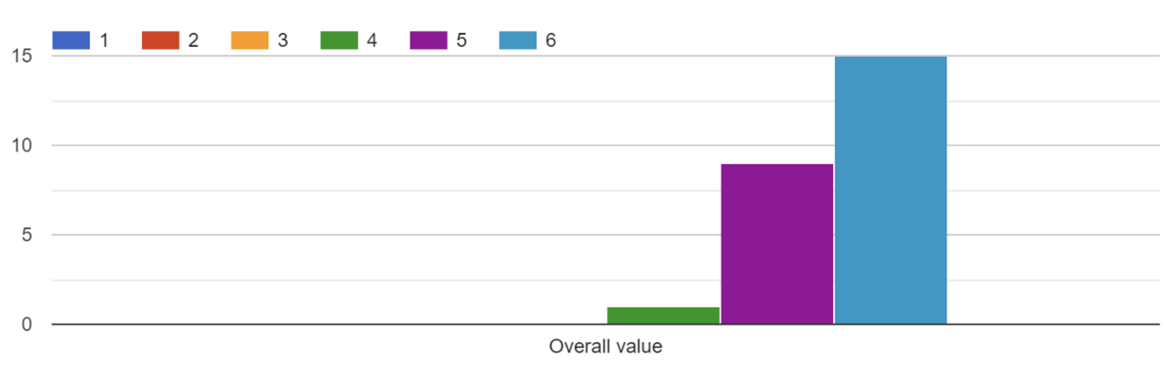


Figure 6.5

Stacked bar chart:

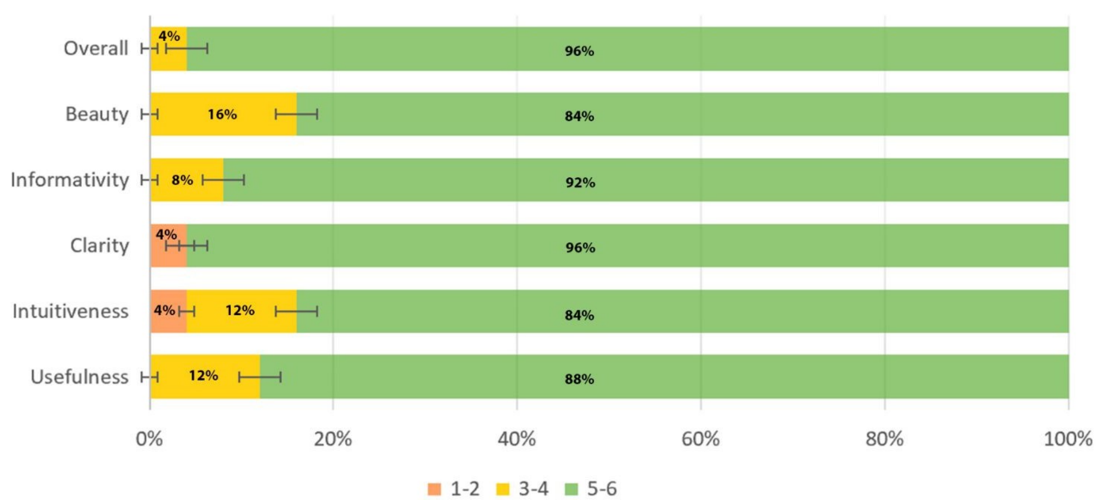


Figure 6.6

The graph depicts the findings of the survey. Very satisfactory results are shown, especially with regard to Informativity and overall value. The sample of subjects returns more than 80 % of the results in the optimal range (5 and 6). The least positive aspects turn out to be beauty and intuitiveness, where there are 16% of answers below 4, while clarity has the highest percentage in the optimal range, 96%.

The following chart is a Correlogram between the variables used in the Psychometric Questionnaire:

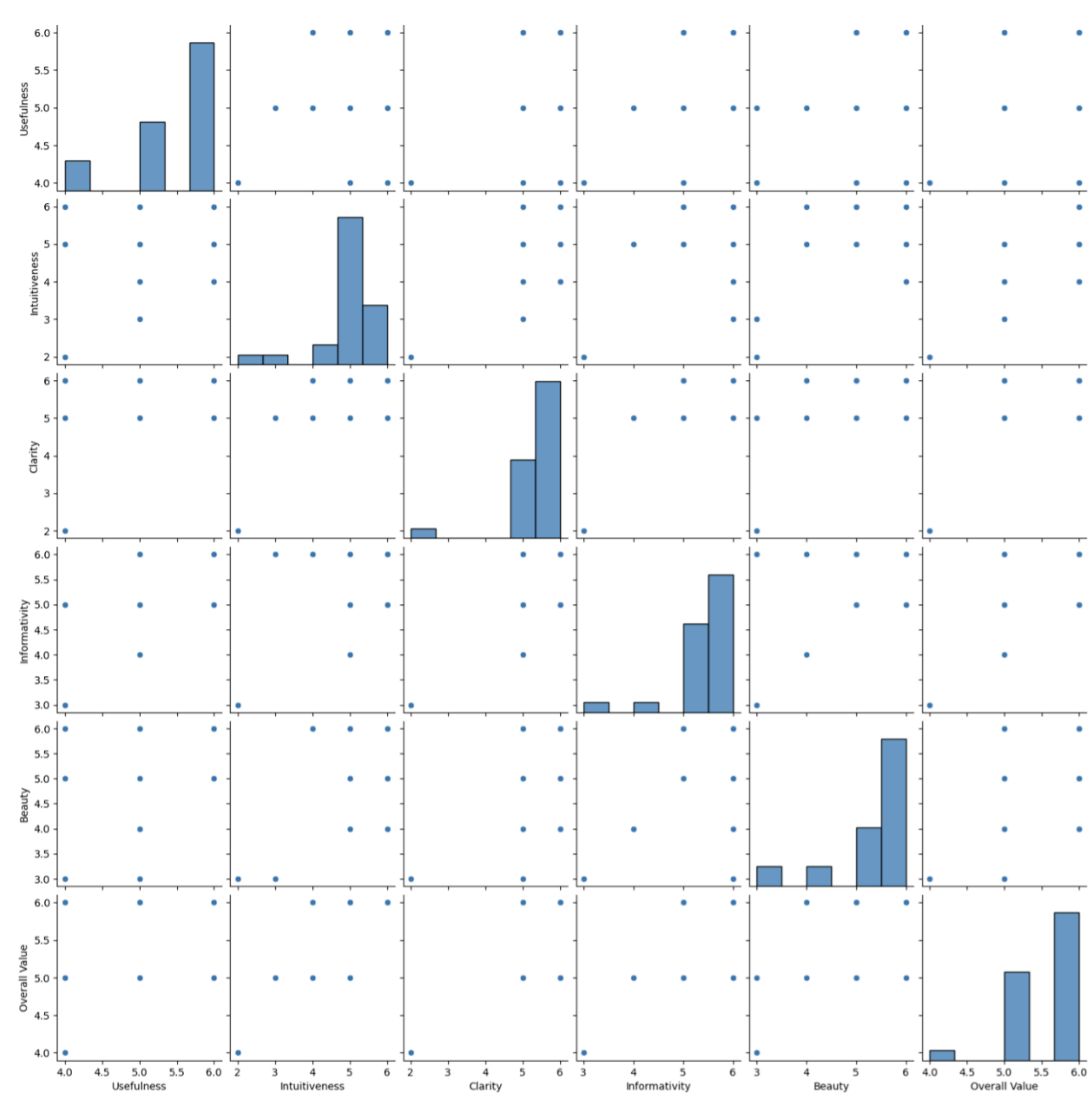


Figure 6.7

Chapter 7

User Test

The last evaluation that we have done is the user test. It is a summative assessment; therefore this test was performed at the end of the story construction. We decided in advance 3 different tasks to be solved by the users.

1. The overall difference of the total number of passengers from 2005 to 2016 is bigger in the category “Low fare” or “Other”?
2. In which of the two plots the total percentage of the flights is more homogeneously distributed?
3. In March, which is the continent with the percentage difference higher than the previous month?

After that, we studied the user behavior, we tried to detect errors, difficulties and misunderstandings during the tasks execution. In addition, we measured the time used for completing each task by each user. In order to not have biased results, we changed the order of questions, so we didn't risk having longer time in the last task due to fatigue. To perform the test, two of us were involved each time in order to support the user with an initial presentation (facilitator) and to observe the user and to record his behavior (observer). This allowed us to collect results for 11 users that will be reported into a table.

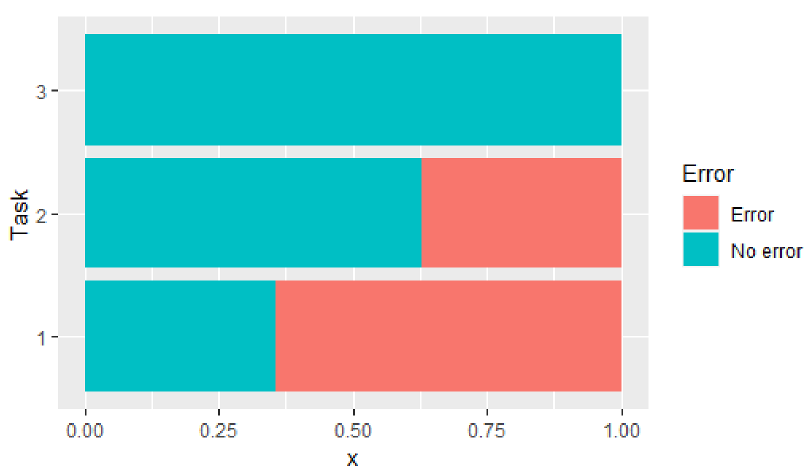


Figure 7.1: Errors in each task

As we can see in figure 7.1, users have some troubles in performing the first task. We can make a differentiation between experts and non-experts. We should report the thing that, for the last ones, the box plot is not an easy graph. Moreover, the task required to use one of the filters on the right. Although these points, the thing that brought non-experts to make errors was the fact that the y-axis was not fixed. It changes the range of values depending on whether we work with “Low fare” (800k-100k) or “Other” (400k-200k). As a consequence, “Low Fare” had the higher expansion visually, but, if users look carefully to the y-axis, they can notice that in “Other” the difference between the high and the low whisker is bigger! On the other hand, the experts gave the correct answer.

The tasks number two and three were easier and we recorded low timing in both categories (see figure 7.2) and few errors. In particular, the second task is related to the fifth plot of the story where we compare domestic and international flights. It is interesting to notice that the users arrived to the solution in different ways: for someone, color conveys information, instead someone else read the percentages on the screen. Also for the third task, users took different paths to reach the goal. In general they searched for the bigger bubble and then they read the percentages in them to be sure of the answer. It is curious to notice that no one used the shades of color to be lead to the solution.

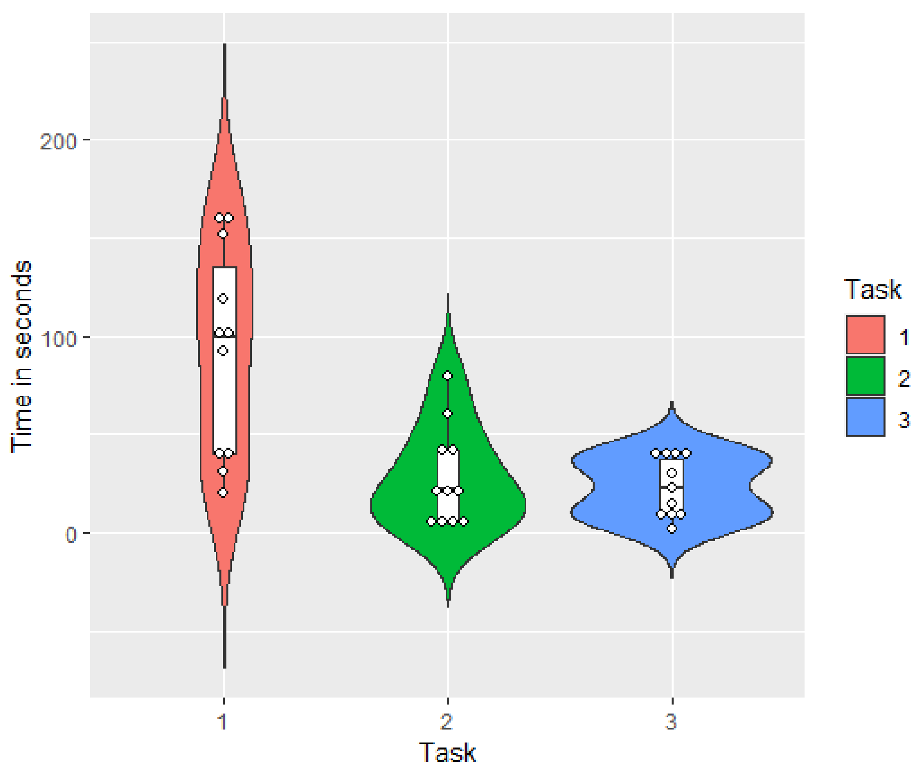


Figure 7.2: Violin plot of executing time of each task

In order to check for efficiency in the user test, a t-test can be applied comparing the times of execution of each task. There is no statistically significant difference between the averages completion times of these three tasks.

Task	level experience	execution time	task accomplished	n.errors
User 1				
1	non-expert	100 s	no	1
2	non-expert	7 s	yes	0
3	non-expert	23 s	yes	0
User 2				
1	non-expert	20 s	no	1
2	non-expert	6 s	yes	0
3	non-expert	2 s	yes	0
User 3				
1	non-expert	40 s	no	1
2	non-expert	20 s	yes	0
3	non-expert	8 s	yes	0
User 4				
1	non-expert	31 s	no	1
2	non-expert	4 s	yes	0
3	non-expert	9 s	yes	0
User 5				
1	non-expert	152 s	yes	1
2	non-expert	6 s	yes	0
3	non-expert	38 s	yes	0
User 6				
1	expert	40 s	yes	0
2	expert	41 s	yes	0
3	expert	15 s	yes	0
User 7				
1	expert	160 s	yes	0
2	expert	60 s	no	1
3	expert	10 s	yes	0
User 8				
1	non-expert	160 s	yes	0
2	non-expert	80 s	no	1
3	non-expert	40 s	yes	0
User 9				
1	non-expert	92 s	no	1
2	non-expert	43 s	no	1
3	non-expert	38 s	yes	0
User 10				
1	non-expert	104 s	no	1
2	non-expert	23 s	yes	0
3	non-expert	42 s	yes	0
User 11				
1	expert	119 s	yes	0
2	non-expert	22 s	yes	0
3	non-expert	30 s	yes	0

Chapter 8

Conclusions and Future Developments

We can state that the infographic we produced was well appreciated. It received positive responses about the majority of the aspects from the contributors that we interviewed, who found the content interesting and, in particular, they emphasised the informativity and usefulness of our design.

In the previous chapters, we have presented and discussed several examples of data visualizations, learning common pitfalls and helpful tricks along the way. As we have seen, developing a simple and effective data visualization is a complex process, and this analysis left us satisfied for the team-effort we invested in it. We believe that this project allowed us to learn on a professional level how to deal with a new tool such as Tableau, but above all to delve into a field that was new to us, the data processing and data representation, and we consider ourselves very pleased about that.

For anyone wishing to resume and continue our analysis, we report some possible changes that could be added in the future, derived from the people who were interviewed and the results of the user test. Indeed, being the latter a summative test, and therefore carried out at the end of the design, it did not allow us to modify the design during the project, but rather leave room for future modifications. We learnt that, while it was easy and straightforward for experienced users to obtain the information we desired, for novice users some requests were more difficult to complete, in particular it proved to be challenging to distinguish between changes when there are two different charts to compare. Therefore, we would like to suggest few adjustments: in the first display regarding the difference between estimated and actual trend, overlapping the two graphs would be more helpful in discerning the actual discrepancies. Similarly, the first question from the user test was universally found to be more complex, so we believe that we should not put the filter either with low fare or other, but rather put a filter for "splitting" them, so that in the initial plot there will be both the fares, and by dividing them we can see the two in the same graph. Since we cannot fix the y-axis because otherwise the low fares would be just a small part of the representation that would not be easily identifiable, representing either of them together we believe is the most effective solution. Finally, we report some suggestions we received from the users as well, such as segmenting the third graph concerning the focus on the number of flights by regions instead of by cities, in order to make the density of flights from one country to another more visible without focusing more specifically on a particular airport. As this is a current and evolving topic, we also believe that it would be intriguing, in a post-COVID scenario over the upcoming years, to explore the actual evolution in flight trends: if the growth will continue and how it will take place, whether it will follow the previous pattern or COVID has permanently changed the growth of air travels around the world.

Bibliography

[1] ICAO, The United Nations, “Economic Impacts of COVID-19 on Civil Aviation” ,*Economic Aspects of Covid-19 Aviation*, https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx.

[2] ICAO, United Nations, *Aviation benefits report*, 2019,https://www.icao.int/sustainability/Documents/AVIATION-BENEFITS-2019-web.pdf.

[3] Wikipedia contributors, D., “Aviation — Wikipedia, The Free Encyclopedia”, 2022, https://en.wikipedia.org/w/index.php?title=Aviation&oldid=1128134540

[4] ICAO, Uniting Aviation,“Aviation Benefits: contributing to global economic prosperity”, Feb 28, 2018, https://unitingaviation.com/news/economic-development/aviation-benefits-for-a-better-future/

[5] ACI, airports Council International,“The impact of COVID-19 on the airport business—and the path to recovery”, February 24, 2022, https://aci.aero/2022/02/24/the-impact-of-covid-19-on-the-airport-business-/

[6] Xavier Olive, ,“Impact of COVID-19 on worldwide aviation”, 2022, https://traffic-viz.github.io/gallery/covid19.html

[7] Wikipedia contributors, ,“List of busiest airports by passenger traffic — Wikipedia, The Free Encyclopedia”, 2022, https://en.wikipedia.org/w/index.php?title=List_of_busiest_airports_by_passenger_traffic&oldid=1123552567

[8] Data Browser of Eurostat,“Air transport of passengers by country (yearly data)”, last update: 03/01/2023 11:00 , https://ec.europa.eu/eurostat/databrowser/view/ttr00012/default/table?lang=en

[9] EDA — Air Traffic Data, “Air Traffic Passenger Dataset”, 2022, https://www.kaggle.com/code/lostinworlds/eda-air-traffic-data

[10] Active Conclusion with Credit: Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic and Matthias Wilhelm. ”Bringing Up OpenSky: A Large-scale ADS-B Sensor Network for Research”. In Proceedings of the 13th IEEE/ACM International Symposium on Information Processing in Sensor Networks (IPSN), pages 83-94, April 2014, “COVID19 Air Traffic Data”, 2022, https://github.com/ActiveConclusion/COVID19_AirTraffic

[11] Angela Locoro and Federico Cabitza and Rossana Actis-Grosso and Carlo Batini, “Static and interactive infographics in daily tasks: A value-in-use and quality of interaction user study”, *Computers in Human Behavior, vol. 71, pg. 240-257, 2017*, issn: 0747-5632, doi: https://doi.org/10.1016/j.chb.2017.01.032, url: https://www.sciencedirect.com/science/article/pii/S0747563217300389, keywords: infographics, Human-data interaction, Usability study, Information-interaction-design quality model.

[12] Tableau Public, https://public.tableau.com/app/discover, last update: 2022.