Logo, company name

Description automatically generated

**FIRST VISUALIZATION PROJECT**

**(TABLEAU)**

**ISM 6361**

**Data Visualization**

Professor Ryan LaBrie

Ngoc Oanh Nguyen

Winter, 2023

1. **State a business reason for selecting your tools (problem you would like to solve).**

Air travel is one of the most popular methods of transportation in the world, particularly for passengers traveling long distances. It is considered as the safest means of transportation in which a person can engage.

“Aviation accidents” are not solely confined to planes crashing. This category includes other aircraft such as helicopters, ultra-lights, gliders, etc. According to statistical reviews from the National Transportation Safety Board (NTSB), plane crashes and other aviation accidents have steadily decreased since 2001. The total number of U.S. general aviation accidents went from 1,728 to 1,085 in 2020. There were only five fatal accidents among 32.2 million flights in 2022.

While trends of risk of flying seem to be continuing downward steadily to exceptionally low levels, that does not mean it is risk-free. The injuries and number of fatalities resulting from aviation accidents unfortunately tend to be far greater than other means of transportation.

Using statistics to identify the root cause of these aviation accidents can help prevent more in the future. Therefore, in this project, I will use the Aviation Accident Database & Synopses up to 2023 to carefully analyze the trends that are emerging to improve the quality and safety of traveling by airplane.

1. **Document how/where you got your data (if it is publicly available, or internal for a work project).**

This Aviation Accident Database & Synopses, up to 2023 is an available dataset on Kaggle ([Dataset Link](https://www.kaggle.com/datasets/khsamaha/aviation-accident-database-synopses?datasetId=521&sortBy=voteCount)).

It is the NTSB (National Transportation Safety Board) aviation accident database, which contains information from 1962 and later about civil aviation accidents and selected incidents within the United States, its territories, and possessions, and in international waters. It is displayed as a preliminary report is available online within a few days of an accident. Factual information is added when available, and when the investigation is completed, the preliminary report is replaced with a final description of the accident and its probable cause.

The dataset may be useful for researchers, aviation engineers, and others interested in identifying the root cause of these accidents and can help prevent more for future research and practice in the aviation field.

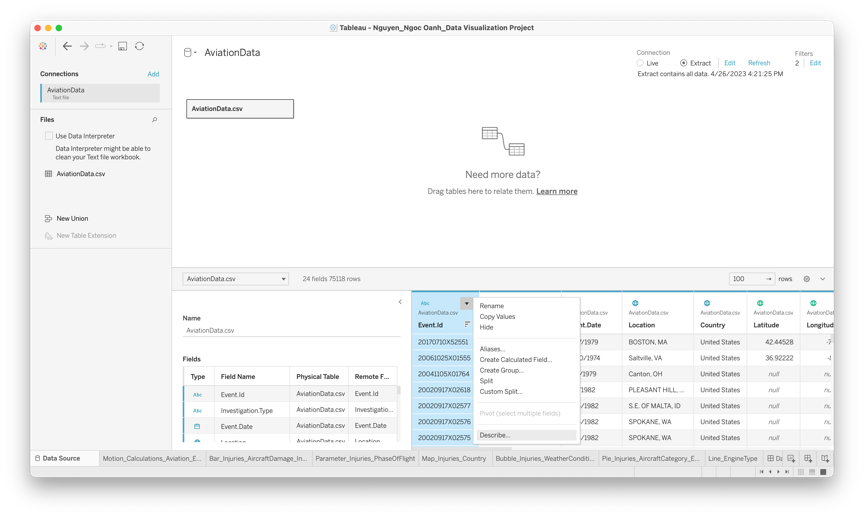
1. **Document how you used the tool. Many tools are super rich in features, and you probably won’t be exploring them all, explain the parts you did use.**

Firstly, I have to load the dataset named “AviationData” which is a csv file into the tool using the “More…” option in “To a File” widget.

Graphical user interface, application

Description automatically generated

After the dataset is loaded, you can select which attributes needed to create the visualizations you want. You can rename, hide, custom split data, change the data type of attributes according to your desire. I removed some attributes that are complicated to visualize or with distinct number, which are most in the string data type including Report.Status, Registration.Number, etc.

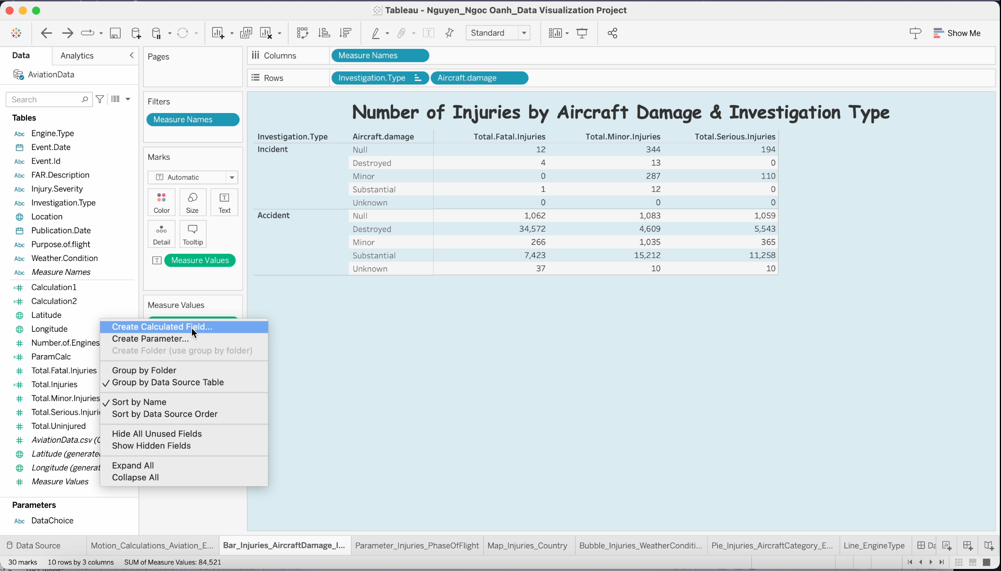


I also using “Filters” to only keep non-null values of Publication.Date attribute and Event.Date one, so I could be able to visual measures continuously without any disruptions from missing values. I also directly extract data for convenience for uploading the story to the public site and faster data processing.

Graphical user interface, text, application

Description automatically generated

Then, I create a calculated field by right-clicking in the measures’ pane (in green color).



The measure “Total.Injuries” was made as the formula below.

**Graphical user interface, text, application, Word

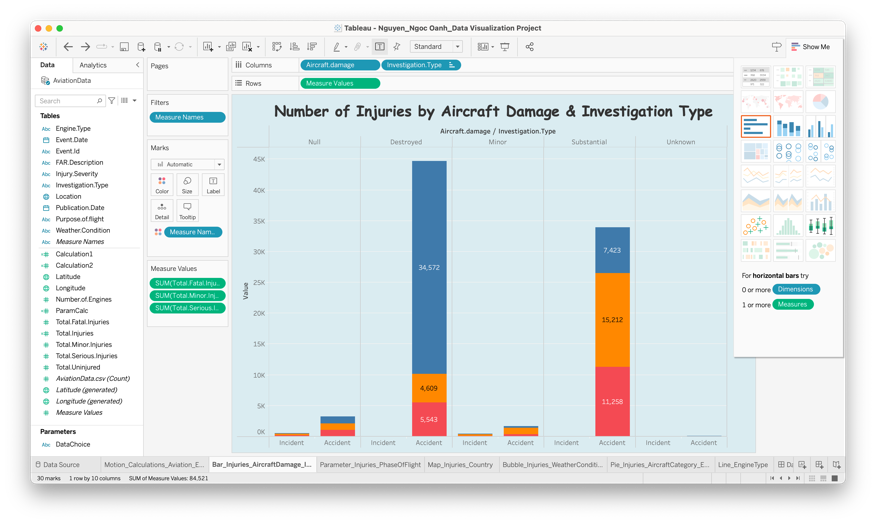
Description automatically generated**

A worksheet is created by clicking Worksheet/New Worksheet, which contains a separate visualization. By putting the attributes and measures into the Columns and Rows field, a basic table with corresponding order will be created.

Graphical user interface, table

Description automatically generated with medium confidence

By using the “Show me” pane in the right corner, you could choose the chart type like bar, line, bubble, etc., which is appropriate to the characteristics of data in order to express the implication.



You can add motion to make the animation of a chart, which is often popular for line charts. You have to change the “Marks” from “Automatic” to “Circle” or “Square”, and show “All” in “Marks to show history for” and “Trails”. As a result, you could see how the lines move over the time period.

Graphical user interface

Description automatically generated

In a parameter, you can choose what measure of the chart by directly switching the pane.

A picture containing chart

Description automatically generated

To create a parameter, you have to right-click the “Parameters” pane, then set up the allowable values you wanted. After that, you create another calculated field with corresponding formulas as follows.Graphical user interface, application

Description automatically generatedGraphical user interface, text, application

Description automatically generated

A map could present a location with measures based on longitude and latitude.

Graphical user interface, application

Description automatically generated

In a dashboard, a lot of related sheets or images can be put in that. It is created by clicking Dashboard/New Dashboard.

Graphical user interface

Description automatically generated

Graphical user interface, application

Description automatically generated

You can create an action in a dashboard (Dashboard/Actions).

Graphical user interface, table

Description automatically generated

You can filter or highlight the same data from different charts by clicking on it, going to the sheet by clicking on the chart, or adding a hyperlink so you can see that site below the chart when hovering.

Graphical user interface, application, Word, website

Description automatically generated

A story basically is a set of sheets and dashboards. It is created by clicking Story/New Story.

Graphical user interface, text, application

Description automatically generated

1. **Explain why you chose which visualizations/charts.**

First and foremost, I choose the chart type based on the nature of the data.

When I want to show how the number of injured people changes over time, I use line graphs combined with motion to create animations. That will help with better visualization.

To show the severity of injuries, including fatal, serious, and minor, I chose a stacked bar chart so that each of these types of casualties can be compared. Furthermore, I also incorporated the Aircraft damage type to emphasize the extent of the casualty types on a case-by-case basis.

To represent the number of casualties based on the broad phase of flight, bar graphs are used, and they are also sorted in ascending order to show how many people were injured in each phase of flight. This is a parameter chart so the measure can be customized, based on the degree of injury, including fatal, serious, minor, and uninjured.

The map is used when I want to show the number of people injured by each country. Figures will show when I hover over that place.

To show the link between Weather Condition and Purpose of flight in terms of casualties, I chose a bubble chart.

To show which Aircraft category accounts for the largest number of plane crashes and the proportion of each Engine type based on the Aircraft Category, I used the proportional pie charts.

To represent the top 5 engine type in Aviation data, I used the area chart. It helps me compare the data of each engine type with each other.

1. **Give an explanation/analysis of the output. What did you learn or uncover?**

Chart, line chart

Description automatically generated

As can be seen from the line chart above, the reciprocating engine accounted for the largest number of crashed aircraft, possibly because it was the most common engine in civil aircraft during the study period. Therefore, it is difficult to conclude that there are more bugs happening inside this engine than other types. Other types of engines have remained at a low level over the years.

Chart, bar chart

Description automatically generated

We can see that most people get fatal injuries when the plane is destroyed. And when the plane has significant damage, the number of people with minor injuries is more than those with fatal and serious injuries. However, these numbers are still very large, demonstrating the serious damage of aviation accidents to human lives. Air travel can be considered the safest means of transport, but once an accident occurs, the consequences will be extremely serious.

Chart, bar chart

Description automatically generated

The above parameter graph of the number of people injured based on each phase of flight can vary the measure of severity such as fatal, serious, minor, and uninjured. The number of people injured in the null phase is always the highest of all measures, which proves that it is very difficult to collect information about plane crashes. Maybe those are accidents so catastrophic that the information black box cannot be found. Next, we can conclude that the takeoff, maneuvering, and cruise phases cause the most serious damage to human life. Therefore, those responsible in the aviation industry such as pilots, mechanical engineers, and logistics staff need to work together and improve technology and equipment to minimize problems that can occur in the air. this stage, in order to avoid accidents.

Map

Description automatically generated

Based on the colors displayed in the map above, we can see that the number of people injured in plane crashes in the US accounts for the majority. However, this data set is mainly in the US, some surrounding areas, and flights depart and end there, so data comparisons between countries are disproportionate and there are certain limitations.

Chart, bubble chart

Description automatically generated

The bubble chart shows the correlation between weather conditions and the purpose of flight in the number of injuries. VMC is Visual Meteorological Conditions in which pilots have sufficient visibility to fly the aircraft maintaining visual separation from terrain and other aircraft, which basically equals good conditions for flying. IMC is Instrument Meteorological Conditions, which require the use of instruments, typically cloudy or low visibility. The biggest bubble is VMC-Personal, which means most injuries are from personal flight purposes, typically traveling on a commercial plane, in good condition. Additionally, there are a lot of green bubbles, which indicates that airplane accidents are not totally due to the weather but other reasons. However, bad weather could increase the number of injuries to some extent.

Graphical user interface, chart, application

Description automatically generated

The pie chart of the number of people injured by aircraft category and engine type shows that the largest pie representing the largest number of casualties belongs to null aircraft. Followed by airplane and helicopter respectively. The proportions of each type of engine are also revealed through the magnitude of the angles.

1. **Conclude with the 3 W’s (What Went Well, What Did NOT go Well, What Would you do Differently Next Time).**

* What Went Well

The data preparation did not take so much time for me since the dataset is quite a clean one, so I did not have to cleanse them that much. Everything is relatively in good condition which facilitates me to focus on manipulating the dataset.

Tableau is extremely rich in features and visualizations, thus there is a lot of things to.

Tutorials provided by the Professor are very detailed, which made it a lot easier for me to get used to this new tool and also to manipulate the data.

* What Did NOT go Well

First, finding a dataset with all required attributes to be able to create calculations, map with location, line chart with motion, etc. is not easy at all but extremely time-consuming. In the first place, I chose another dataset but that one is not cleaned with a lot of missing values which are important. Therefore, I had to switch to other datasets before deciding to deal with this Aviation Accidents dataset. However, there is still a lot of missing data and confusingly classified attributes in this dataset so I have to considerably eliminate the unnecessary one.

In addition, there are many attributes that are string types such as Report.Status or Registration.Number, which can be difficult to illustrate with other metrics. Besides, this dataset does not have many numeric attributes, which makes it difficult for me to freely choose measures to display on the graph. I only have the aviation accidents count and total of injuries with their severity items including fatal, serious, minor, and uninjured. This caused me to repeatedly use each of the injured people variables over and over. Because of the limited data, my charts are not quite as beautiful and consistent as I would like. Sometimes the performance is quite forced without really showing the meaning.

* What Would I do Differently Next Time

I want to spend more time finding the data to understand each attribute more comprehensively. In particular, I had to carefully consider the amount of numeric data to be able to perform calculations and visualize the data more easily.

I also want to try a different larger dataset next time with different visualization tools like Grafana or FusionCharts to broaden my knowledge in this field.

1. **The hyperlink (web address) to your Tableau Public Site.**

<https://public.tableau.com/app/profile/ngoc.oanh.nguyen/viz/Nguyen_NgocOanh_DataVisualizationProject/Story_CivilAviationAccidents>