



Aviation Accident Data Visualizer

Yetao Chen, Yifan Xu, Tianyi Zhang



Background

Aviation safety is always an important research area in transportation safety.

#1 Although aviation accident is one of the rarest transportation accidents, the fatal rate of aviation accident is among the highest.

#2 Causes of aviation accidents are extremely complicated and unpredictable.

In last 2 years, Boeing 737 MAX has suffered 2 hull loss accidents, including Lion Air Flight 610 crash in 2018, and Ethiopian Airlines Flight 302 crash in 2019. We hope this project can give people some intuitive feelings about aviation accidents in U.S., and arouse our awareness of aviation safety as well.



Background

This project is intended to create a tool for users who are interested in aviation accident in U.S.. The basic functions of this package are data query and visualization, which can be accomplished with pandas, matplotlib and folium in Python.

Data

The data we use is from DATA.GOV. It is named Aviation Data and Documentation from the NTSB Accident Database System and released by National Transportation Safety Board.

InvestigationType	AccidentNumber	EventDate	Location	Country	Latitude	Longitude	AirportCode	AirportName
Accident	CEN20LA015	10/27/2019	White Bear Township, MN	United States	45.116667	-92.996111	6MN9	BENSON
Accident	GAA20CA051	10/26/2019	Tavares, FL	United States	28.800556	-81.727778	FA1	Tavares
Accident	CEN20LA014	10/24/2019	Swea City, IA	United States	43.384166	-94.320278	N/A	
Accident	CEN20FA012B	10/23/2019	Hebbronville, TX	United States	27.372500	-98.633334	N/A	
Accident	CEN20FA012A	10/23/2019	Hebbronville, TX	United States	27.372500	-98.633334	N/A	
Accident	CEN20LA013	10/22/2019	Lamesa, TX	United States	32.735000	-101.743333		N/A
Accident	GAA20CA044	10/21/2019	Brooksville, FL	United States	28.473611	-82.455556	BKV	Brooksville-Tampa Bay Rgnl
Accident	GAA20CA041	10/20/2019	Muskeget Island, MA	United States	41.334722	-70.299444	MA55	Muskeget Island
Accident	GAA20CA040	10/20/2019	Reno, NV	United States				N/A
Accident	GAA20CA038	10/20/2019	Melbourne, FL	United States	28.102778	-80.645277		N/A
Accident	CEN20FA009	10/20/2019	White Plains, AL	United States	33.749722	-85.723334		N/A
Accident	GAA20CA037	10/19/2019	Albuquerque, NM	United States	35.084444	-106.650278		N/A
Accident	GAA20CA036	10/18/2019	Marble Canyon, AZ	United States	36.810833	-111.644444	L41	Marble Canyon

This is a small part of view of the data set opened in Excel. We can see not all of the accidents have their accurate geographic locations, and 90% of the accidents in the data set are in the US.



Use cases

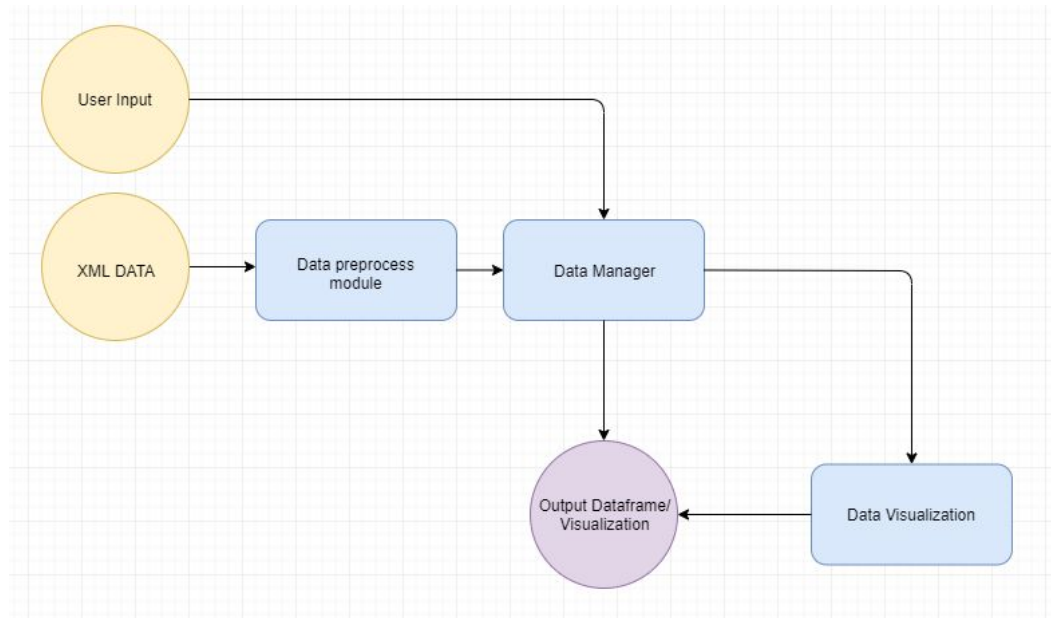
- Extract clean dataframe
 - Actor: Package users
 - Preconditions: Preprocess module imported
 - Flow: The user calls 'dataset()' function from Preprocess module. The module takes 'AviationData.xml' and clean the data with functions in the module and returns a dataframe.
- Multiple condition querying
 - Actor: Package users
 - Preconditions: DataManagement module imported, dataframe input
 - Flow: The user calls 'column_query()' function with proper dataframe and conditions as input. The module takes the dataframe and condition strings and returns a subsection of the original dataframe.



Use cases

- Visualization for data correlation (Single Graph)
 - User is able to selectively plot different data entry processed from Data Management module
 - Often used for determining relationship between different data sets such as time vs. number accident.
 - Offering intuitional results for entry level users or complex datasets.
- Visualization for data comparison (Multiple Graph)
 - Available for user to select in order to indicate the correlations across different properties of data set.
 - Excellent for comparing data entries by plotting on the same plot with same scale.
 - Very flexible in data selection without defining anything in advance.

Structure Diagram

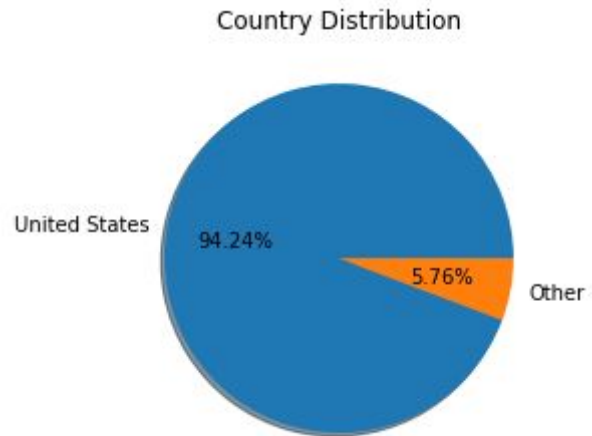
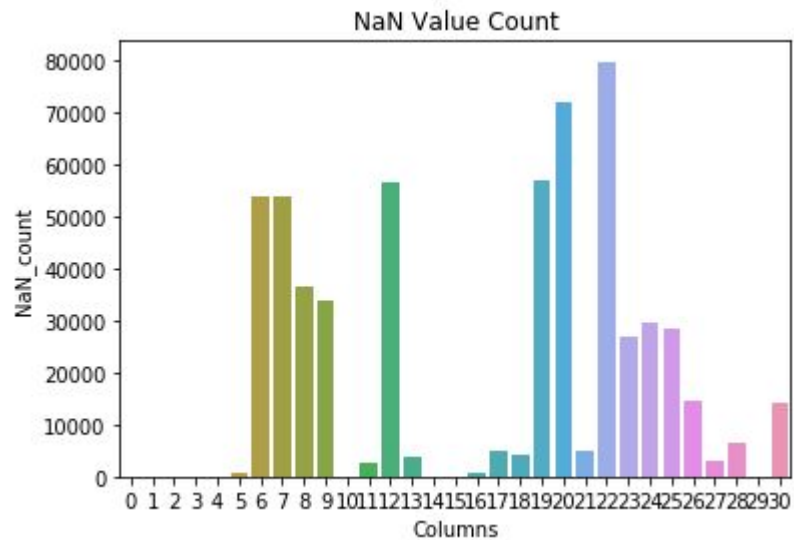




Data Preprocessing

- Problems to deal with:
 - NaN values
 - Missing values
 - Pool distribution on features
 - Feature separation

Data Preprocessing





Data Manager

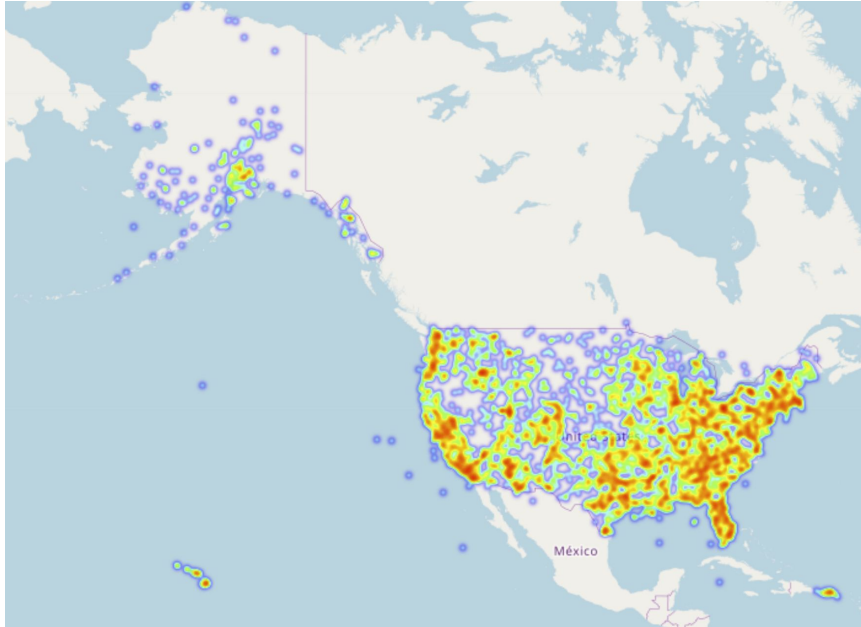
- Data Manager
 - Manage queries from user input
 - Process data upon request
 - Return a subset of original dataset



Data Visualization

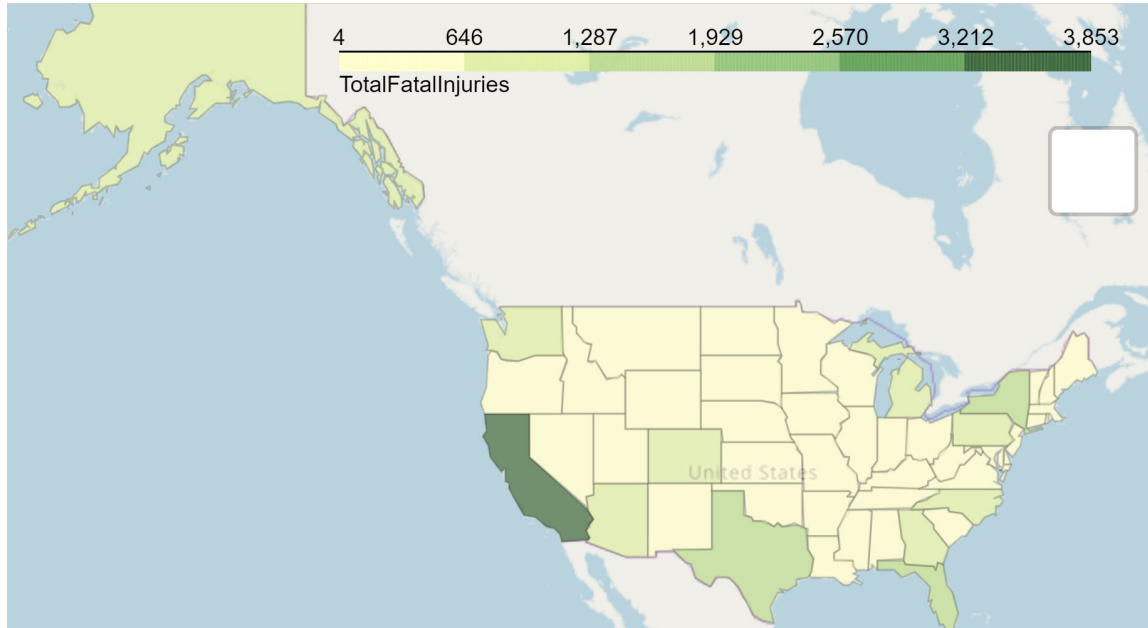
- Data Visualization:
 - Take processed data as input from Data Manager
 - Provide visualization output with chosen form(Geographic map, histogram, bar charts, heatmaps, colormaps)
 - Plots methods and data sets defined by user with many choices.

Data Visualization and Demo

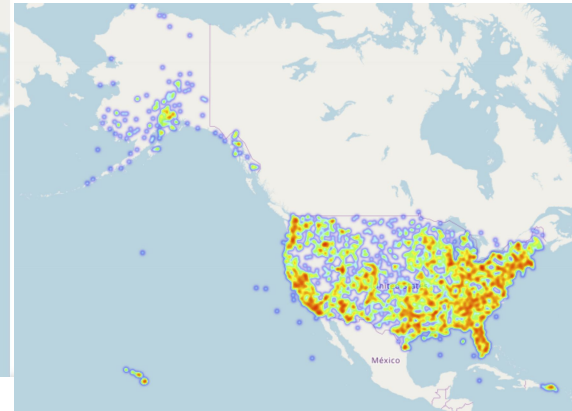


- Visualization example of all aviation accidents from Jan. 1, 1990 to Jan. 1, 2019 in US, in terms of total accident number, regardless of injury.

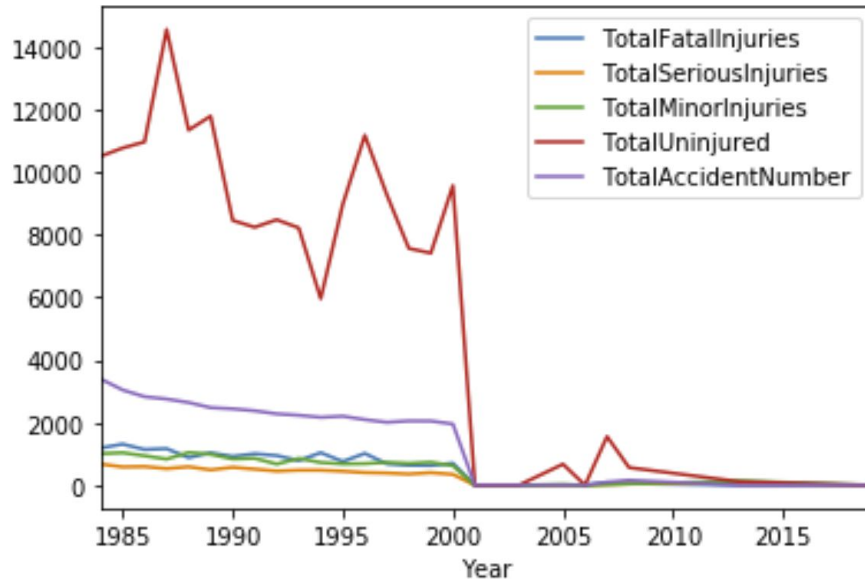
Data Visualization and Demo



- Visualization example of all aviation accidents since establishing of NTSB in US in terms of total fatal injuries.

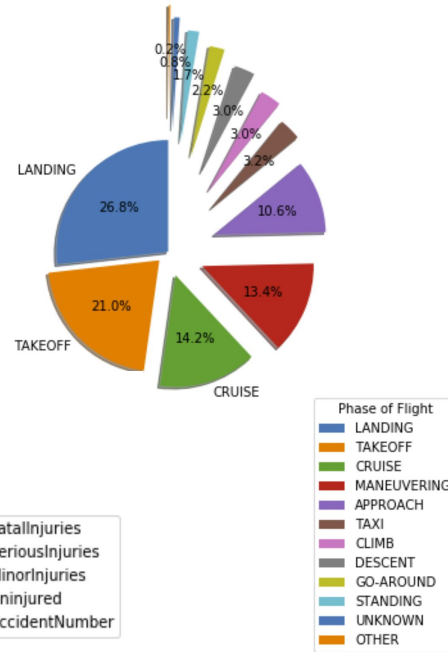
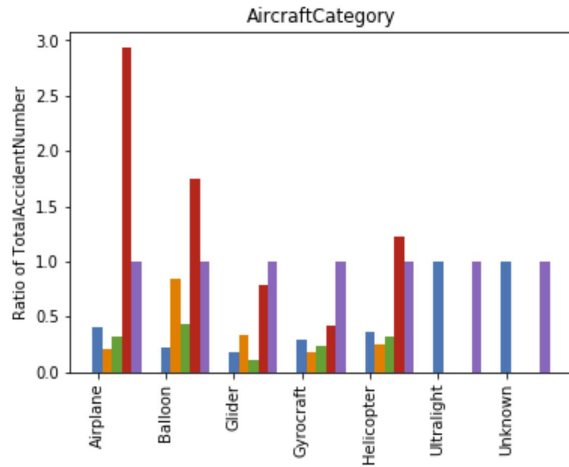


Data Visualization and Demo



- Number of accident and number of passenger involved in past 30 years.

Data Visualization



- Left: Normalized injury rate in different type of aircraft.
- Right: Accident distribution during different flight hase



Project Structure (Github)

```
AviationAccidentDataVisualizer/  
  |- AviationAccidentDataVisualizer/  
    |- Modules/  
      |- init .py  
      |- Preprocess.py  
      |- DataManagement.py  
      |- test DataManagement.py  
      |- test Preprocess.py  
      |- AviationData.xml  
    |- init__.py  
  |- Examples/  
    |- Examples.py  
  |- docs/  
    |-Component Specification.txt  
    |-Functional Specification.txt  
  |- LICENSE  
  |- README.md  
  |- setup.py  
  |- Requirements.txt
```




Lessons learned

- **Valuable Programming Experience:**
 - Basic coding in Python; Documentation style; Dataframe management in Python.
- **Learned and Used Several Python Libraries:**
 - Pandas for dataframe management; Matplotlib and Altair for visualization; Folium for map visualization.
- **Software Design in Real World:**
 - Defining user cases; Specifying components and function; Package structure.
- **Working with Git:**
 - Working in teams; Command line; Version control.



Future work

- Additional visualization features
- Dataset expansion
- User Interface



Thank You