Downloaded data from https://data.world/datamil/vietnam-war-thor-data

Data was split by year and each year had several tables

Used python script by Jie Jenn to combine all tables within a year into one table in sql server https://www.youtube.com/watch?v=r9Ry3CBXXW0

Created empty tables using the column names of the tables in sql server, then ran Jie Jenn's script to populate. Created one table per year

Had to clean data a little as the script was running into the error "bulk load data conversion (truncation)" for a particular column in the tables. Turns out the script was recognizing colons in the column as delimiters. The column was titled "Additional Info" and generally followed the format "mission: alpha-tango-ur mum:123". Removed this data since it was not relevant to the analysis

Brainstorming ideas for visualizations:

- [x] # of total missions flown by country (bar)
    - [ ] Sorting by kinetic vs. non-kinetic missions
- [x] Bombings over time (by date), month-year, by country (line)
    - [ ] Sorting by kinetic vs. non-kinetic missions
- [x] Operations by day on a map, filter by country (geo)
    - [ ] Sorting by kinetic vs. non-kinetic missions
- [x] Number of missions flown by branch over the entire war (bubble)
    - [ ] Sorting by kinetic vs. non-kinetic missions
- [ ] ~~Military branch pie chart (hehe xd)~~
- [ ] ~~Mission count by time of day (00:00-24:00), filter by mission function ie strike armed recon… (bubbles along time, scatterplot)~~
- [x] Aircraft type (tree map) (avg mission duration for each aircraft)
    - [x] Instead maybe we use the PERIOD OF DAY column and list number of missions for each part of the day (ie. Day: 1700, Midday: 145, Night: 78) (bubble chart description)
    - [ ] Sorting by kinetic vs. non-kinetic missions
    - [ ] ~~Do calculation between Time on target and time off target (avg mission duration?), don't include values where time on and time off are zero~~
        - [ ] ~~Had to drop this one since time values were all messed up :(~~
- [ ] Operation the mission was apart of
- [ ] ~~Cloud cover (maybe link up with time?)~~

Standardized dates using CONVERT(date, ___)

Checked for duplicates among the converted dates using CTE's and window function ROW_NUMBER

Compared different aircraft columns to see what the difference was. Some aircrafts that flew 1 mission were difficult to search for on google so I wrote a query to select for the aircrafts that flew more than 1 mission AKA the relevant aircrafts. Make sure to mention this in the dashboard!

The timeontarget and timeofftarget are strings. Must convert into 24 hour time format. Did some colon insertion into the varchar to make it recognizable to the datetime function. Assumed that values that are only 2 numbers long are in minutes

> UPDATE: time values are riddled with errors and inconsistencies. Even after formatting several of the values are nonsensical and ambiguous. Not gonna do this one.

Created new column labelled Lat_Long combining longitude and latitude for all tables

Created new column labelled MSNDATE_Converted which standardizes dates

Created new column labelled PERIODOFDAY_Standardized which converts all evening missions to night and morning missions to day

Created new column labelled milservice_formatted, turns all nulls into "other" and "USA" to "USAF" (made an educated guess that USA was probably a typo for USAF since there were significantly fewer entries for USA than USAF)

Returned to the geo map and added a label that counts the number of times a single location was hit on the same date. This was done to prevent adding all 3 million+ rows to the tableau dashboard, instead, we plot all the unique locations (for each day) and display the number of times it was hit

Another issue with the geo map visualization: A lot of nonsensical locations are plotted on the map (ie. kinetic missions in china, saudi arabia, etc.). According to research, the main countries associated with aerial missions in the vietnam war include Laos, Cambodia, Vietnam and Thailand as well the neighboring waters of these countries. I will set longitude and latitude limits for the data points as I am almost certain there were no incursions outside the indochina area. Although it is still likely there are data input errors in the indochina region, we'll cut down on the most obvious ones.

- Westernmost limit: Mae Sariang District, Thailand (18.548030, 97.369644)
- Northernmost limit: Lung Cu, Vietnam (23.390527, 105.322781)
- Easternmost limit: South China Sea (15.344668, 117.196773)
- Southernmost limit: South China Sea (6.051594, 107.924745)