

Data Appendix

Our final dataset, Cleaned Aviation Data Final.csv, reflects the data cleaning from our original dataset, Aviation Data 2015-2025.csv. The final dataset includes 29 columns with each row representing a single aviation accident or incident and its corresponding data. The following variables from our final dataset were used in our analysis script.

Variables Used in Cleaned Aviation Data Final.csv:

- **EventDate**

- *Definition:* Contains the date of the aviation accident in the format of YYYY-MM-DD.
- *Missing values:* 12682(0)

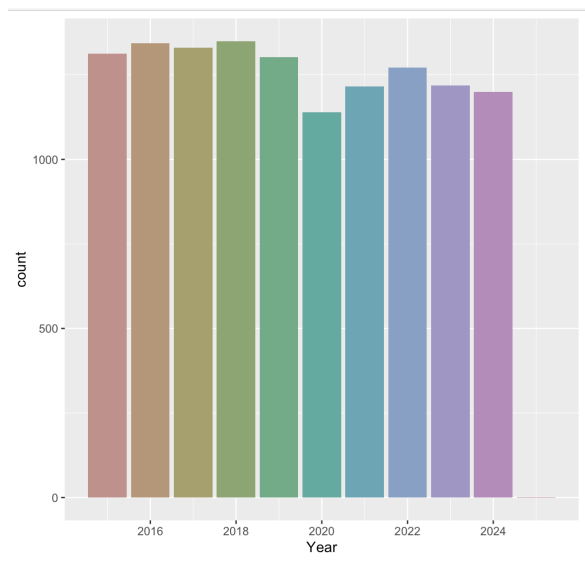
- **Year**

- *Definition:* Contains the year (2015 - 2025) of the aviation accident.
- *Missing values:* 12682(0)
- *Frequency Table:*

```
> table(Data$Year)
```

2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
1312	1343	1330	1350	1303	1139	1216	1271	1218	1199	1

- *Frequency Distribution:*



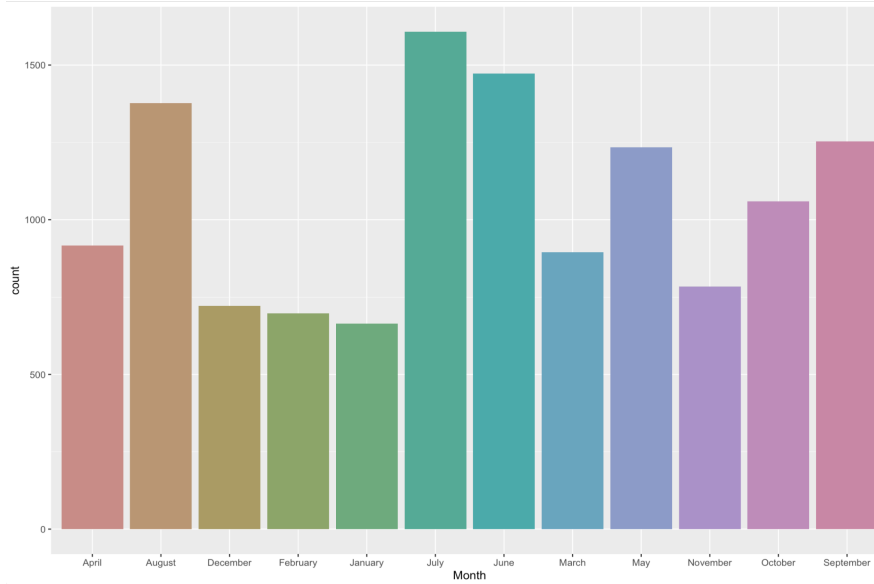
- **Month**

- *Definition:* Contains the month of the aviation accident in numerical format.
- *Missing values:* 12682(0)
- *Frequency Table:*

```
> table(Data$Month)
```

April	August	December	February	January	July	June	March	May	November	October	September
917	1377	722	697	664	1608	1472	895	1234	784	1059	1253

○ Frequency Distribution



● FatalInjuryCount

- *Definition:* Contains the number of fatal injuries from the accident.
- *Missing values:* 12682(0)
- *Summary Statistics:*

```
> summary(Data$FatalInjuryCount)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.0000	0.2801	0.0000	16.0000

○ Frequency Distribution

Distribution of Fatal Injury Count

