Intro and Purpose

This analysis was done by Cassidy Bell, Neelam Prasad, Aaron Suarez, and Tyler Beringer. We analyzed the Kaggle dataset named **US Tornado Dataset 1950-2021.** Our goal was to explore historical tornado events.

Data and Data Cleaning

The dataset contains 67,558 entries with information on things like magnitude, injuries, and fatalities. To better uncover insights about trends and patterns, we narrowed in on specific parts of the data.

We noticed some problems with the dataset. Types had to be changed and columns had to be renamed. Plus, duplicates and magnitudes equal to -9 had to be removed.

Changing the type of date:

```
df['date'] = pd.to_datetime(df['date'])
```

Renaming the columns:

```
df = df.rename(columns={ 'yr': 'year', 'mo': 'month', 'dy': 'day', 'date': 'date', 'st':
'state', 'mag': 'tornado_magnitude', 'inj': 'injuries', 'fat': 'fatalities', 'slat':
'start_latitude', 'slon': 'start_longitude', 'elat': 'end_latitude', 'elon':
'end_longitude', 'len': 'tornado_length', 'wid': 'tornado_width' })
```

Removing the duplicates:

```
duplicates_to_remove = df[df.duplicated(keep='first')]
df = df.drop_duplicates(keep='first')
```

Removing the magnitudes equal to -9:

```
magnitudes_to_remove = df[df['tornado_magnitude'] == -9]
df = df.drop(magnitudes_to_remove.index)
```

By the time we concluded the data cleaning process, we had a quality dataset that was ready for use. It was time to start answering questions.

Research Questions

[...]

Limitations and Biases

One of the limitations that our group came up with was when there were multiple tornadoes that touched down in a similar area, on the same day, with the same magnitude.
[Biases]

Conclusion

This was an interesting analysis which revealed unique insights about US tornadoes.
[insights]