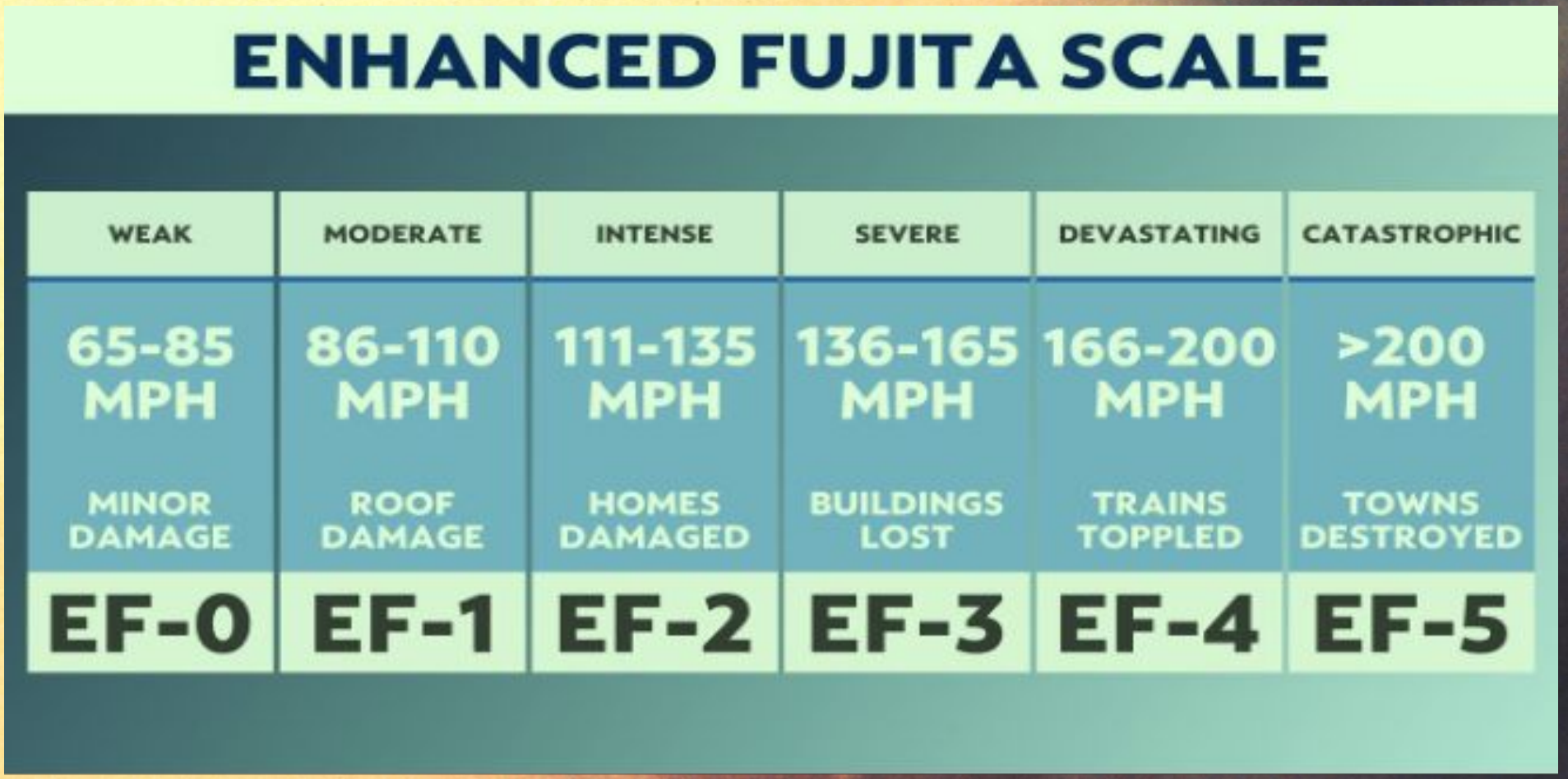
US Tornados (1950-2022)

Team 1:
Fabi Estrada
Uzor Francis
Ally Blitch
Ivette Reese


Top 10 Close
TORNADOES

Introduction



Data Engineering

```
# Check data type and missing values
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 68693 entries, 0 to 68692
Data columns (total 27 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   om            68693 non-null  int64
 1   yr            68693 non-null  int64
 2   mo            68693 non-null  int64
 3   dy            68693 non-null  int64
 4   date          68693 non-null  object
 5   time          68693 non-null  object
 6   tz            68693 non-null  object
 7   datetime_utc  68693 non-null  object
 8   st            68693 non-null  object
 9   stf           68693 non-null  int64
 10  mag           67937 non-null  float64
 11  inj           68693 non-null  int64
 12  fat           68693 non-null  int64
 13  loss          41523 non-null  float64
 14  slat          68693 non-null  float64
 15  slon          68693 non-null  float64
 16  elat          68693 non-null  float64
 17  elon          68693 non-null  float64
 18  len           68693 non-null  float64
 19  wid           68693 non-null  int64
 20  ns            68693 non-null  int64
 21  sn            68693 non-null  int64
 22  f1            68693 non-null  int64
 23  f2            68693 non-null  int64
 24  f3            68693 non-null  int64
 25  f4            68693 non-null  int64
 26  fc            68693 non-null  bool
dtypes: bool(1), float64(7), int64(14), object(5)
memory usage: 13.7+ MB
```

```
# Check missing values by column
df.isna().sum()
```

```
om                0
yr                0
mo                0
dy                0
date              0
time              0
tz                0
datetime_utc      0
st                0
stf               0
mag             756
inj               0
fat               0
loss          27170
slat              0
slon              0
elat              0
elon              0
len               0
wid               0
ns                0
sn                0
f1                0
f2                0
f3                0
f4                0
fc                0
dtype: int64
```

```
# Drop rows with missing values
df.dropna(inplace=True)
```

```
df['mag'].isnull().sum()
```

```
0
```

```
df['loss'].isnull().sum()
```

```
0
```

```
# Another check missing values by column
df.isna().sum()
```

```
om              0
yr              0
mo              0
dy              0
date            0
time            0
tz              0
datetime_utc    0
st              0
stf             0
mag             0
inj             0
fat             0
loss            0
slat            0
slon            0
elat            0
elon            0
len             0
wid             0
ns              0
sn              0
f1              0
f2              0
f3              0
f4              0
fc              0
dtype: int64
```

Data Engineering cont…

```
df["region"] = [state_regions[x] for x in df.st]
df.head()
```

	om	yr	mo	dy	date	time	tz	datetime_utc	st	stf	...	ns	sn	f1	f2	f3	f4	fc	region	month	num_tornados
0	192	1950	10	1	1950-10-01	21:00:00	America/Chicago	1950-10-02T03:00:00Z	OK	40	...	1	1	25	0	0	0	False	South	October	2499
1	193	1950	10	9	1950-10-09	2:15:00	America/Chicago	1950-10-09T08:15:00Z	NC	37	...	1	1	47	0	0	0	False	Southeast	October	1070
2	195	1950	11	20	1950-11-20	2:20:00	America/Chicago	1950-11-20T08:20:00Z	KY	21	...	1	1	177	0	0	0	False	South	November	904
3	196	1950	11	20	1950-11-20	4:00:00	America/Chicago	1950-11-20T10:00:00Z	KY	21	...	1	1	209	0	0	0	False	South	November	904
4	197	1950	11	20	1950-11-20	7:30:00	America/Chicago	1950-11-20T13:30:00Z	MS	28	...	1	1	101	0	0	0	False	South	November	2209

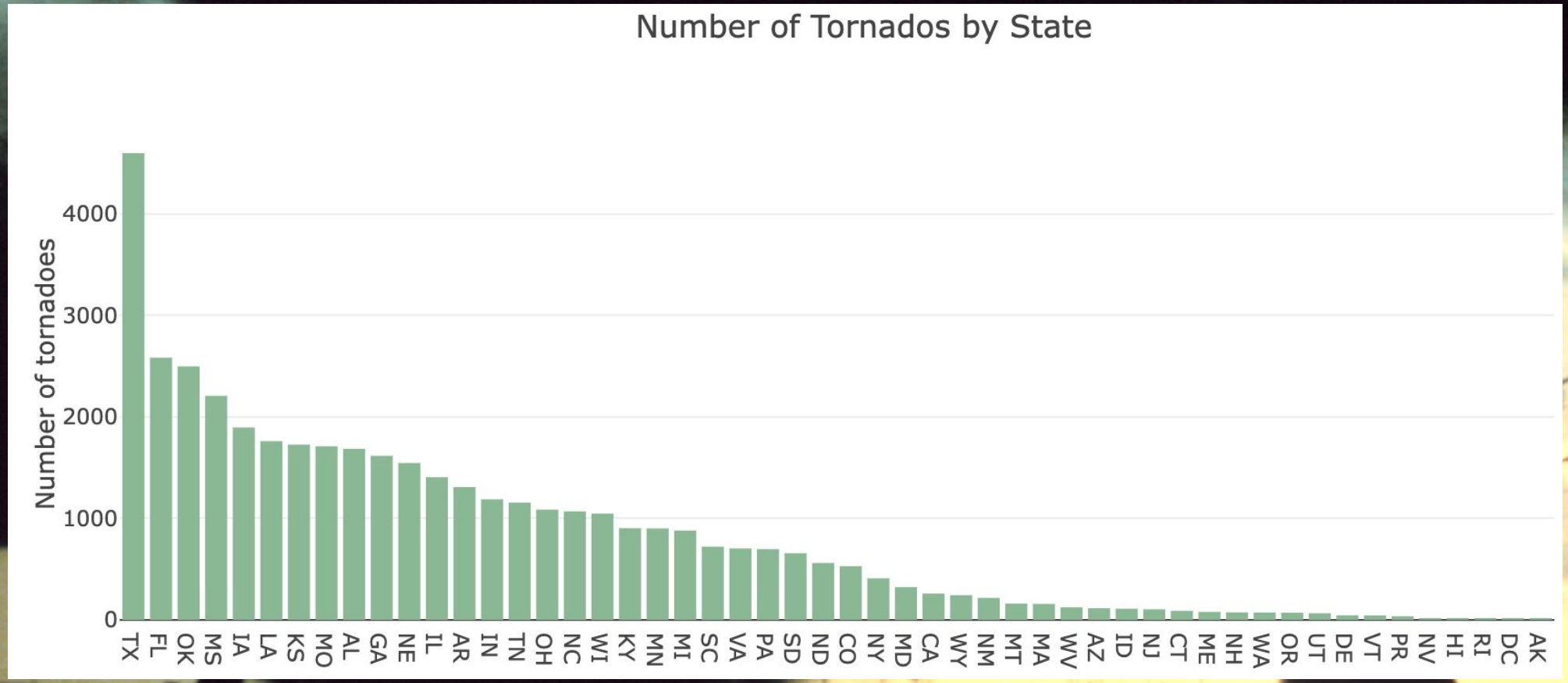
5 rows × 30 columns

Creating the SQLite File

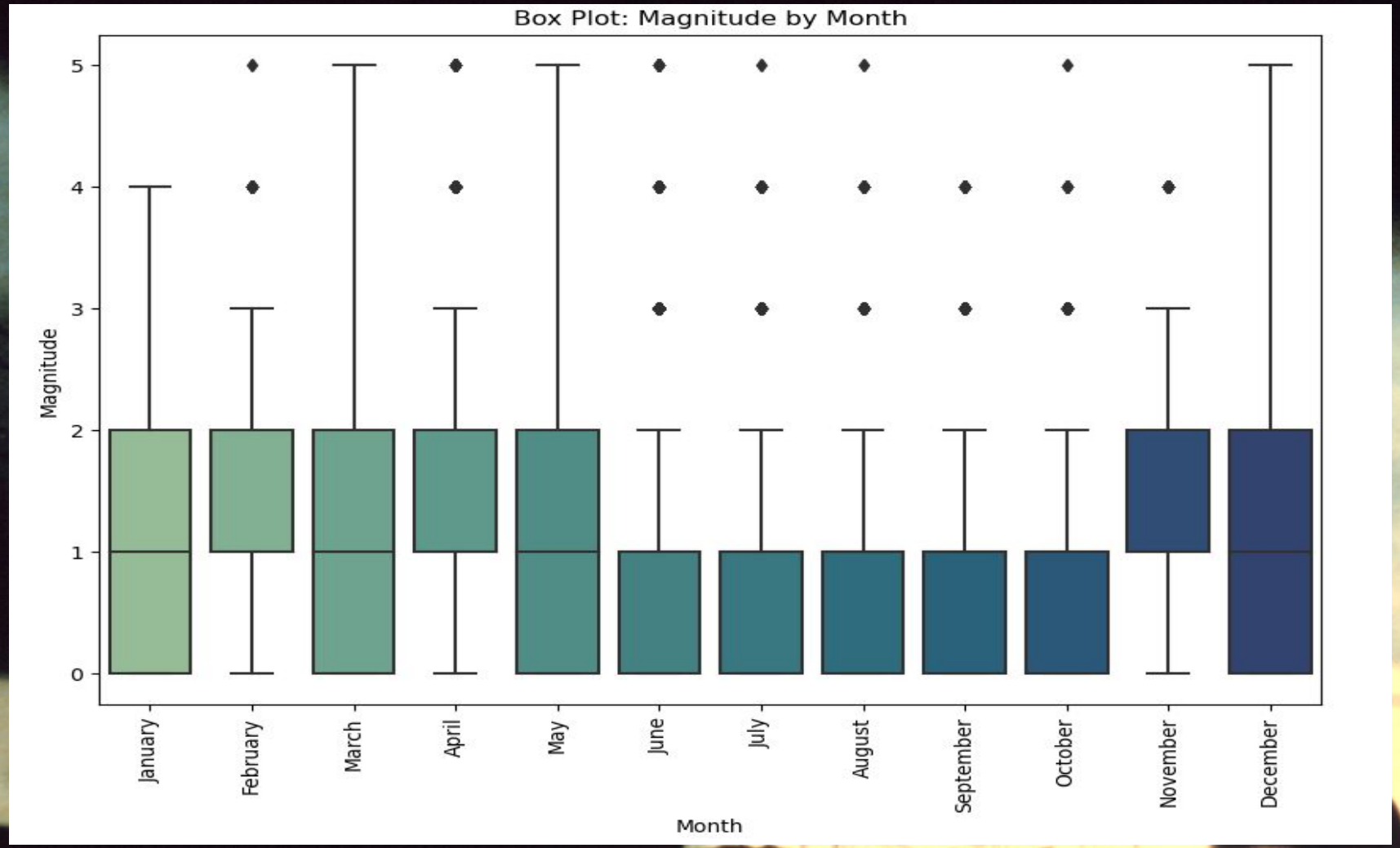
```python
# For sunburst chart allow the user to select ALL or a specific state
if region == "All":
     where_clause = "1=1"
else:
    where_clause = f"region = '{region}'"

query = f"""
    SELECT
        region as label,
        "" as parent,
        count(*) as num_tornados
    FROM
        tornados
    WHERE
        {where_clause}
    GROUP BY
        region

    UNION ALL

    SELECT
        st as label,
        region as parent,
        count(*) as num_tornados
    FROM
        tornados
    WHERE
        {where_clause}
    GROUP BY
        st,
        region;
"""

df_sunburst = pd.read_sql(text(query), con=engine)
data_sunburst = df_sunburst.to_dict(orient="records")

df_sunburst.head()
```

	label	parent	num_tornados
0	Midwest		14617
1	Northeast		2040
2	South		14565
3	Southeast		8407
4	West		1887

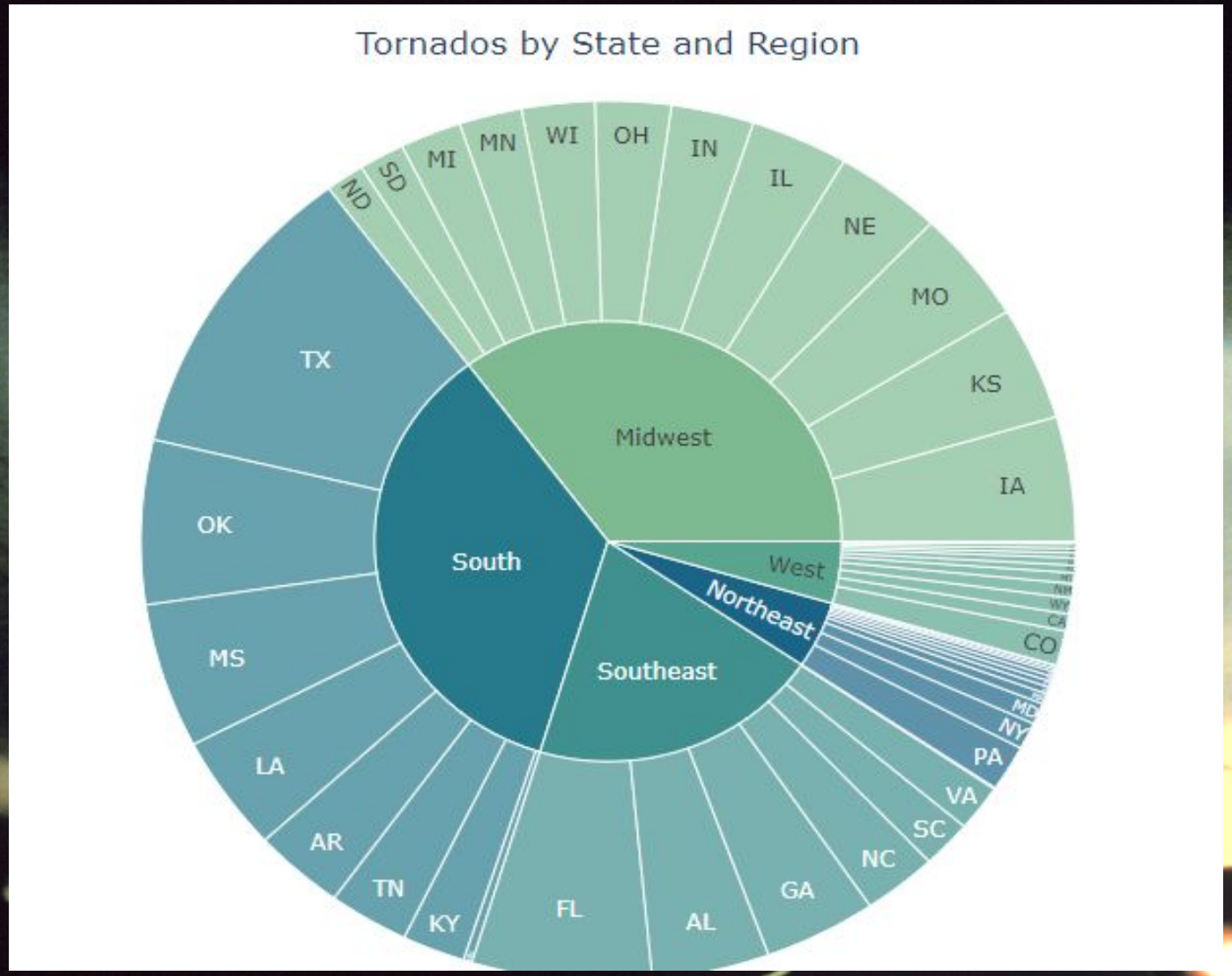
Visualizations (1 of 3)



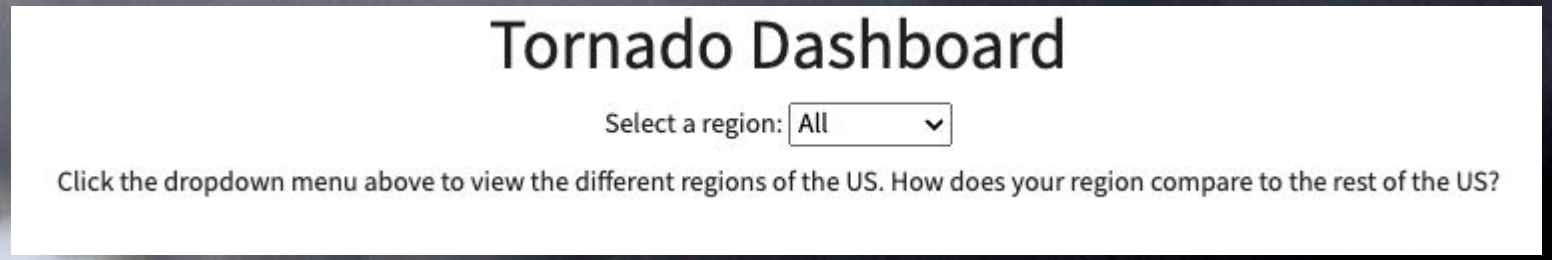
Visualizations (2 of 3)



Visualizations (3 of 3)

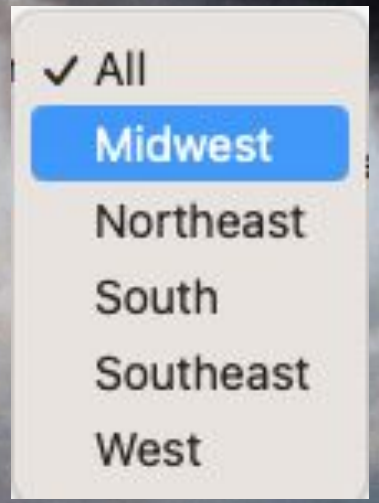


Creating the Dashboard



Using the JavaScript code and Plotly library, we created the dashboard that consists of one dropdown that allows you to filter by region of the US. The dashboard displays the bar chart, box plot chart, sunburst chart, and map. When a dropdown option is chosen, all of the visualizations update accordingly.

The website opens to a home page. The dashboard is the second page on the website, followed by an About Us page.



Findings

1. Between 1950 and 2022 there were 68,693 tornados.
2. The highest number of US tornados occurred in the Midwest with 14,617, closely followed by the South with 14,565.
3. Texas had the highest number of tornadoes with 4,601, while DC and Alaska had the lowest concentration of tornadoes with 2 and 1, respectively.
4. The median magnitude across each month is 2.

Conclusions

1. The magnitudes of US tornadoes in 1950-2022 were typically around 0-2, meaning less severe tornadoes dominate our data.
2. Texas is the most heavily hit by tornados compared to other states.
3. Tornados occur in all US states but most of the tornados occurred in the Midwest region.

Limitations/Bias

Accounting bias – According to the American Meteorological Society, the reporting of tornado occurrence is recorded at the county level. All tornadoes regardless of size/EF rating are counted. However, because of selective reporting weak tornados with minimum impact/losses may go under-reported.

However, there have been some ongoing research on how tornadoes that are likely not be measured can be better captured via supercells technology which is an AI technology.

Future Work

An analysis to determine if climate change affect tornado activity. In what way?

An analysis based on examining the intensity and the path length of each tornado instead of examining number of tornados.

Comparison of starting location and ending location of tornados, instead of only using the starting locations.

Future Work Continued

With the advent of AI and other new technological breakthrough in the 21st century, we believe AI might have a key role to play, one which might be centered on using artificial intelligence(AI) base algorithms to effectively perform regression analysis on Tornado dataset. This good take the shape of both or any of the following two frontiers of AI predictive power;

1. Estimating past tornado occurrence.
2. Predicting future tornado occurrence.

Highlights on estimating past tornado using AI capabilities include:

- Obtaining an AI decision support system for data scientist that they can use to read data.
- Using an AI system which automatically identifies the effective drivers of tornado occurrences in particular regions of the State/Country
- Making tornado dataset more accessible to uses on mobile related devices.

- AI radar-based methods for capturing even the smallest tornadoes that usually get unnoticed according to major reports

Interestingly, with the adoption of AI, we can use automated linear and logistic regression to correctly identify some of our tornado features like duration, speed and area covered.

Work Cited

https://www.kaggle.com/datasets/sujaykapadnis/tornados/data

Horrific EF-5 tornado in Moore, Oklahoma: May 20, 2013 (youtube.com)

https://www.spc.noaa.gov/wcm/data/2008bams.pdf The American Meteorological Society

https://www.gtri.gatech.edu/newsroom/new-approaches-including-artificial-intelligence-could-boost-tornado-prediction

Xpert Learning Assistant via BootcampSpot