Group 4: Evi Storrs, Iris Ding, Mridula Geddam, Amelie Kempf, Shreya Rathi

Group 4 Proposal

The dataset we chose to analyze is the History of Ice Freezing and Thawing on Lake Mendota. The dataset includes information on the days of ice coverage on Lake Mendota from the winter of 1852-1853 to present--how long Lake Mendota froze over each year. Data was collected from the Wisconsin State Climatology Office (SCO), whose goal is to provide reliable weather and climate data for the Wisconsin public. Variables include:

- **Winter:** The corresponding year for the winter. Values are formatted as strings with a hyphen denoting which years the winter corresponds to.
- **Freeze-Over Date:** Which day the lake froze over according to the criteria: *50% of the lake is covered*. Values are formatted as strings with the full month name followed by day number.
- **Thaw Date:** The date when the lake has thawed according to the criteria*: whether one can row a boat between Picnic Point and Maple Bluff*. This criteria was defined to accommodate for 19th century viewers who did not have a high enough vantage point to measure the length and shape of Lake Mendota. Values are formatted as strings with the full month name followed by day number.
- **Days of Ice Cover:** The duration of the lake being frozen in days. Values are integers.

Code to read data:

```python
import pandas as pd

dataframe = pd.read_csv("data.csv")

dataframe.head()
```

	Winter	Freeze-Over Date	Thaw Date	Days of Ice Cover
0	2024-25	January 7	March 15	69.0
1	2024-25	December 25	December 27	NaN
2	2023-24	January 15	February 28	44.0
3	2022-23	December 25	April 2	98.0
4	2021-22	January 7	April 2	85.0

Questions to Investigate:

- How has the amount of days changed over time?
  - Are we seeing shorter periods of freeze over?
- How does having multiple occurrences of freezing affect the total days for a year?
  - Are there longer periods of freeze with multiple occurrences?
- How does the start date impact total days?

Questions for Professor/Consideration:

- Appending data around temperature or humidity or other confounding variables?
- How to format multiple episodes of freezing? As in the screenshot below, 2024-25 has 2 freeze episodes. How should we format them?
-

Winter	Freeze-Ov	Thaw Date	Days of Ice Cover
2024-25	7-Jan	15-Mar	69
2024-25	25-Dec	27-Dec	
2023-24	15-Jan	28-Feb	44
2022-23	25-Dec	2-Apr	98
2021-22	7-Jan	2-Apr	85
2020-21	3-Jan	20-Mar	76
2019-20	12-Jan	22-Mar	70
2018-19	10-Jan	31-Mar	86
2018-19	15-Dec	21-Dec	