

# TITLE OF YOUR PROJECT

Exploring Disparate Data: Part 3 - Final Report

Put your group name here: (anyname that is fancy and funny)

Due November 29th, 2025

List your group members, including their student numbers, here:

- Spencer Robinson (169114936)
- Thomas Sefton (169067096)
- Paolo Spadano (169126042)

## Data Description

### Sea Ice

The data come from NOAA and describe sea ice extent, which is the total area covered by at least 15% ice.

In order to clean the data, we first loaded the two sheets of the .xlsx workbook separately since the data is divided by hemisphere. We then removed unnecessary columns leaving and appropriately naming the month and day columns as well as a column for each year of observation. We then filled in the gaps in the month and pivoted the data so that it no longer had a column for each year. Finally, we sorted the data chronologically by casting the newly formed year column as an integer and converting month to a factor containing the full names of each month.

### Climate Change Awareness

The data come from the Humanitarian Data Exchange and detail awareness of climate change around the world based on surveys.

In order to clean the data, we selected the workbook focusing on aggregate data, pivoting it longer so that countries were no longer each a column. We then converted the answer options on the spreadsheet to more programmatic and clear names (ex. “Refused” became “aware\_refuse”). Finally, we pivoted the data wider to give each level of awareness its own column instead of having a row per country per awareness level.

### Storms

The data come from NOAA (affiliated with the American government) and detail the wind speed, lifespan, category, and location of major storms in the Atlantic and Pacific basins.

In order to clean the data, we read in the two .csv files of interest and separated out the one column containing most of the information using the “,” delimiter. We then replaced any -99 and -999 values with N/A for consistency and created columns for the basin, number, and year (combined), name, and number of entries. We then downfilled the aforementioned columns, filtered out header rows using the fact that they have no

category value, and removed the Entries column (which only had values in header rows). We then separated out the combined BasinNumberYear and time columns. Finally, we renamed the identifier column and cast all numeric variables as such. #TODO

## Combining the Data

Explain how any combinations of data were performed. Explain what kind of join was needed, whether columns had to be modified (for example, matching “country” names.)

## Exploratory Data Analysis

To achieve our goals, we explored the data by looking for historical trends in the development of storms and comparing those to the changes in sea ice year over year. This methodology was chosen in order to identify what exactly was changing about storms and then see if those changes could be attributed to sea ice.

We explored many aspects of the data, but will demonstrate three. These are the changes in hurricane latitude over time, latitude changes as a storm develops, and the relationship between one year’s sea ice and the next year’s hurricane locations.

The first aspect that we found interesting is shown in @ref(fig:insight1). Since observations began in about 1850, the average hurricane has been observed at lower and lower latitudes. This negative correlation has remained consistent despite a few outliers (particularly in 1925).

This insight is supported by the summary statistics in table @ref(tab:summary\_stats)

The next insight that we found is shown in @ref(fig:insight2).

Finally, @ref(fig:insight3) shows ...

## Conclusion and Future Work

Overall, we found <>.

A second paragraph about our findings.

The next steps in this analysis are...

The limitations of this analysis are as follows. (Do not simply list potential issues with sampling, but relate them to your analysis and how they affect your conclusions. An honest and complete acknowledgement of the limitations makes the analysis more trustworthy.)

## References

I am not strict about MLA or APA style or anything like that. For this report, I would much rather have your citations be easy to match to your insights.

The easiest way is to use Rmd’s footnote syntax. This will put a number beside the word where the footnote appears, and the full text of the footnote at the bottom of the page (pdf) or end of the document (html). The syntax is:<sup>1</sup>, where I suggest that you put in something like this<sup>2</sup> to make references for this assignment.

Alternatively, you could make a list of citations with their main arguments and why they’re relevant to your insights, methods, etc.

---

<sup>1</sup>See the source view to see this footnote

<sup>2</sup>The relevance to the insight is . . . From <>, published on <>, url: <>

Average Hurricane Latitude Decreased Over Time  
In the last 150 years

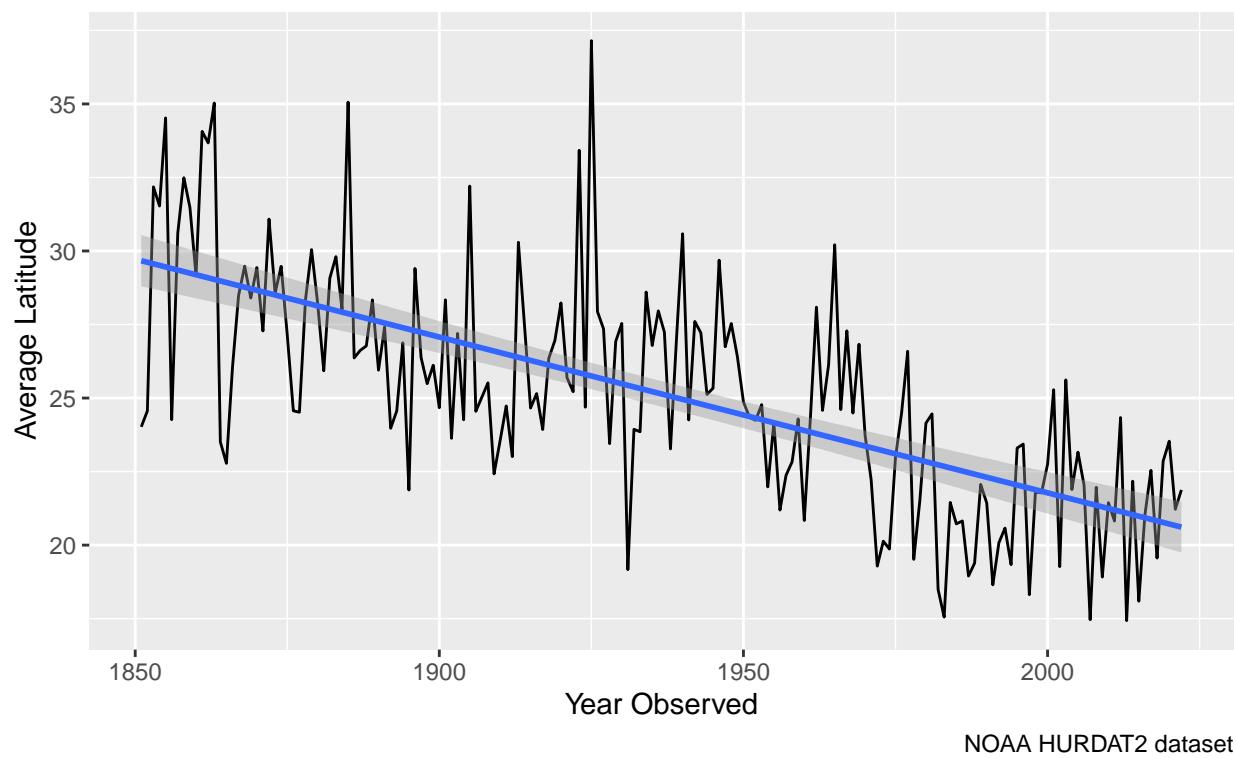


Figure 1: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

## Latitude Increases with Storm Status

Especially in the North Atlantic Basin

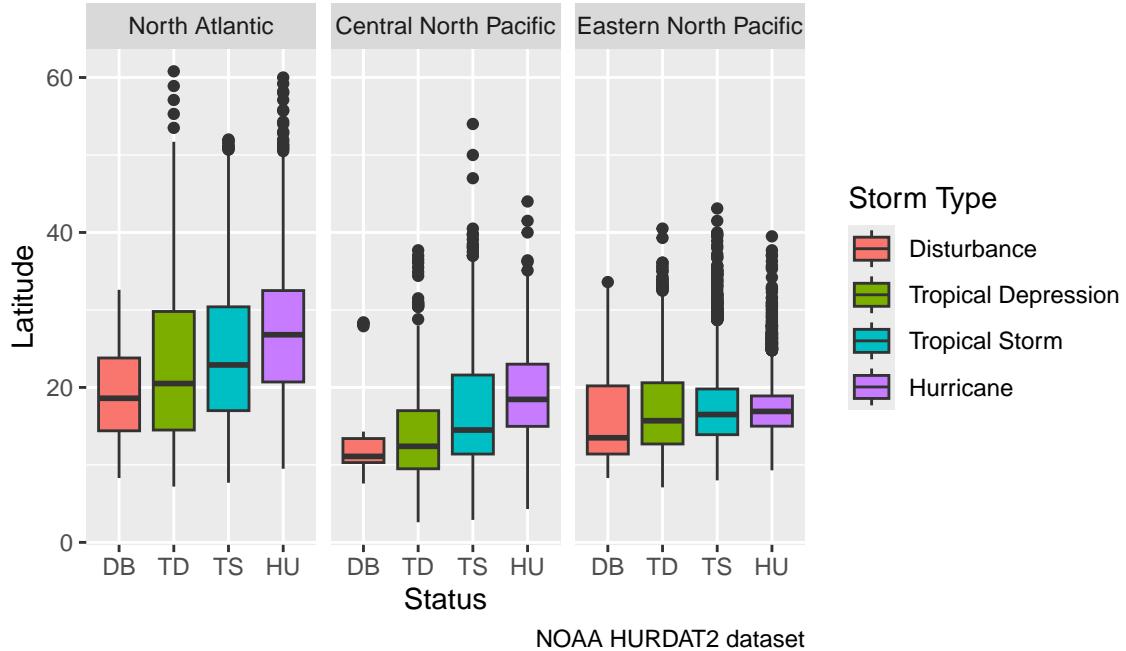


Figure 2: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

The link above also references “bibtex” files. These are also extremely convenient, but have a steep learning curve and they make it difficult to tie them to an insight. If you use bibtext, then make sure that you provide a sentence to describe the source and its relevance when you cite it - don’t just add citations to the end of a sentence (this is common practice in academia, but I want to know that your citations are directly relevant for this assignment).

Hurricane Latitude Inversely Correlated with Previous Year's Sea Ice  
Primarily in the Arctic

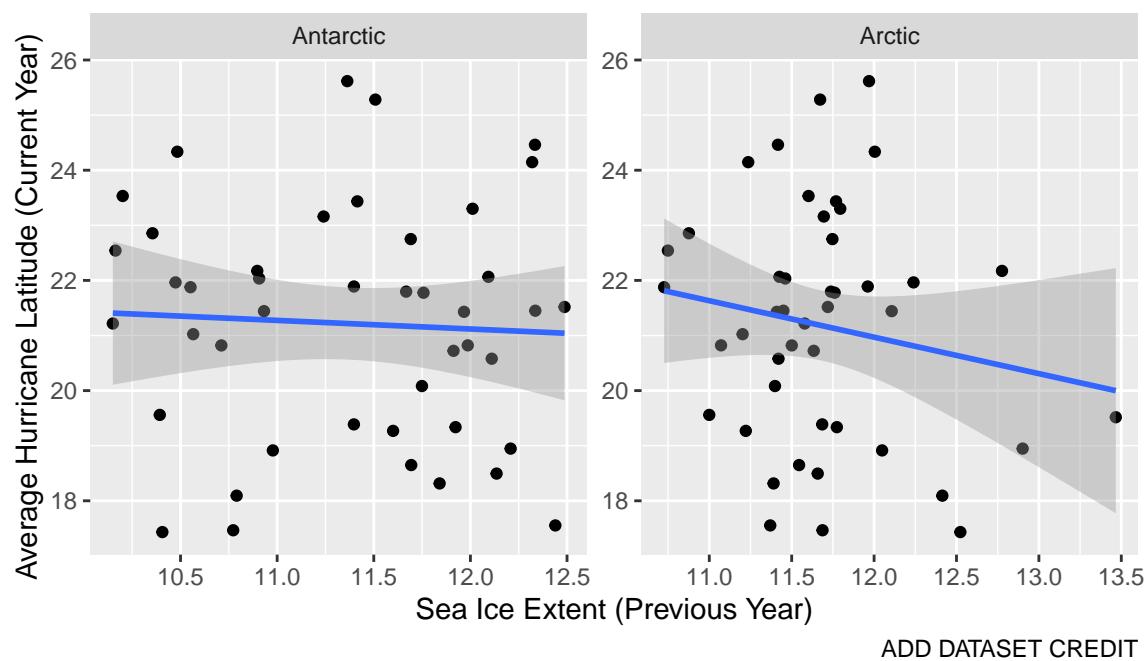


Figure 3: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.