Exploring the Influence of Arctic Sea Ice on Hurricane Latitudinal Shifts and Public Awareness in Affected Countries

Exploring Disparate Data: Part 3 - Final Report

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

Due November 29th, 2025

Group Members:

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

1 Data Description

1.1 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.

1.2 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.

1.3 Tropical Cyclone Exposure

The data come from the Potsdam Institute for Climate Impact Research and contain information about the impact and frequency of storms by country. The data required no cleaning beyond skipping the first six rows of headers.

1.4 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

1.5 Combining the Data

In order to analyze the ice and storm data together both data sets were annualized and then inner-joined using their respective year variables. Although this limits the granularity of the data, it solves the issue of each data set observing only on random days while still allowing year-over-year trends to be visualized.

We also joined the storm count per country with the climate awareness data, which required converting country names to ISO3 format. This was done manually since only the top 10 most affected countries were joined in this way.

2 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 1. 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).

The next insight that we found is shown in 2.

Finally, 3 shows . . .

EXPLAIN TABLE HERE #TODO @ref(tab:summary_stats)

3 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.)

Average Hurricane Latitude Decreased Over Time
In the last 150 years

Median Latitude

Year Observed
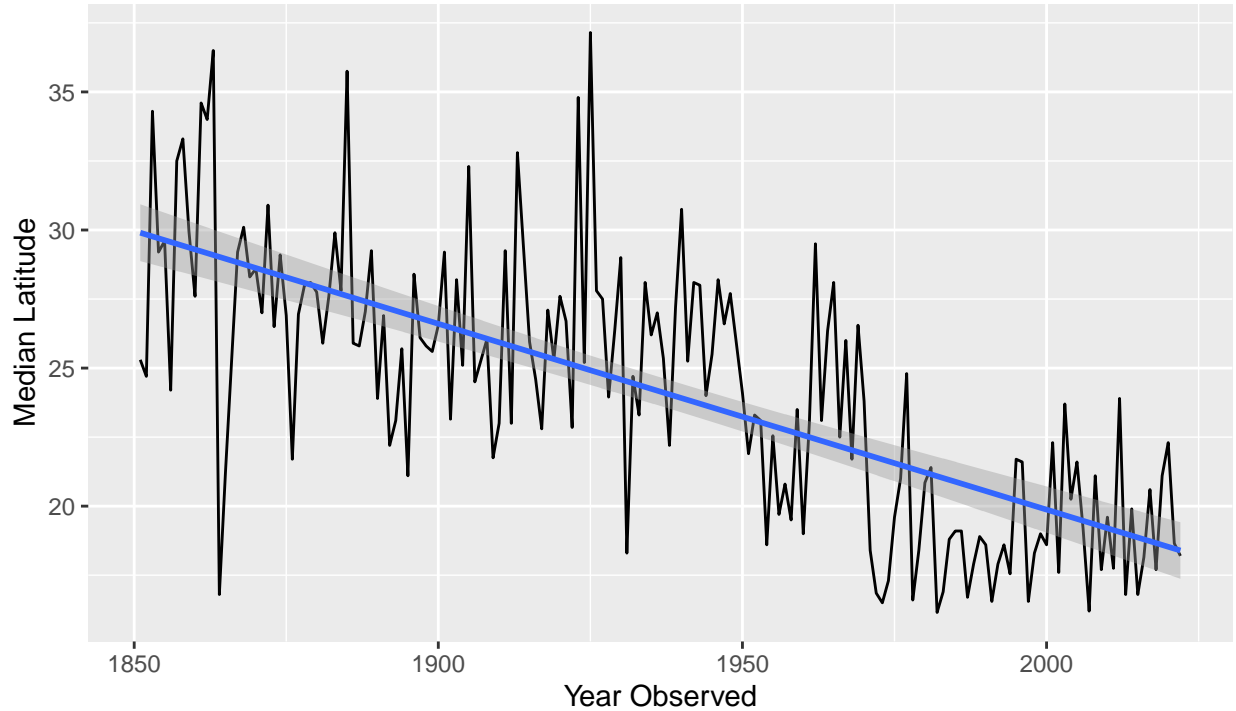
NOAA HURDAT2 dataset

Figure 1: Hurricanes have seen a decrease in average latitude of over 10 degrees since 1850.

country	aware_no	aware_alittle	aware_moderate	aware_alot	storm_count
Philippines	7.36 %	36.79 %	29.4 %	21.78 %	521
Japan	2.26 %	26.56 %	59.65 %	10.9 %	481
Mexico	7.62 %	41.43 %	43.31 %	6.76 %	341
Vietnam	17 %	42.91 %	19.2 %	16.95 %	281
Australia	1.13 %	26.41 %	48.1 %	23.95 %	243
United.States	2.22 %	26.47 %	48.14 %	22.96 %	239
Taiwan	5.94 %	50.74 %	33.88 %	9.08 %	227
India	18.52 %	36.06 %	23.99 %	18.11 %	154
Lao.People's.Democratic.Republic	21.02 %	50.97 %	18.24 %	8.57 %	125
Hong.Kong	7.01 %	51.29 %	31.13 %	10.33 %	108
Worldwide Average	**12.67 %**	**35.83 %**	**35.16 %**	**13.92 %**	

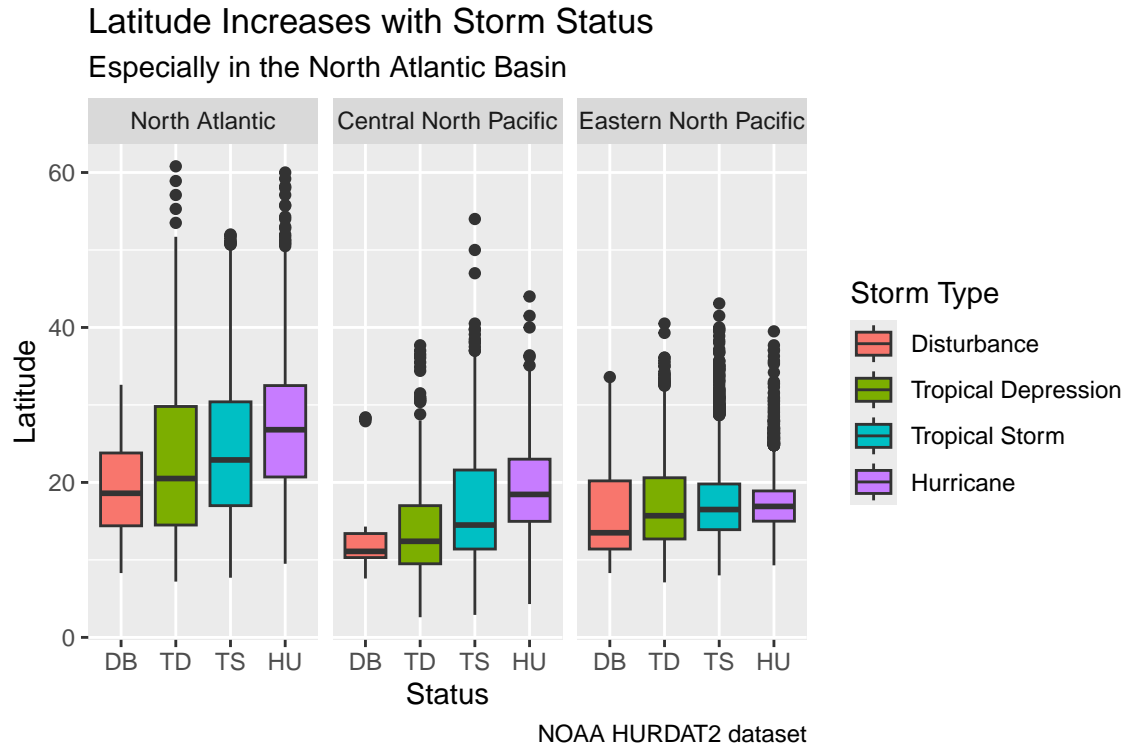Figure 2: This is a figure caption that you will need to change in order to get good marks in the visualization rubric items.

4 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:[1], where I suggest that you put in something like this[2] to make references for this assignment.

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

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 it's 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 assignmnet).

---

[1] See the source view to see this footnote

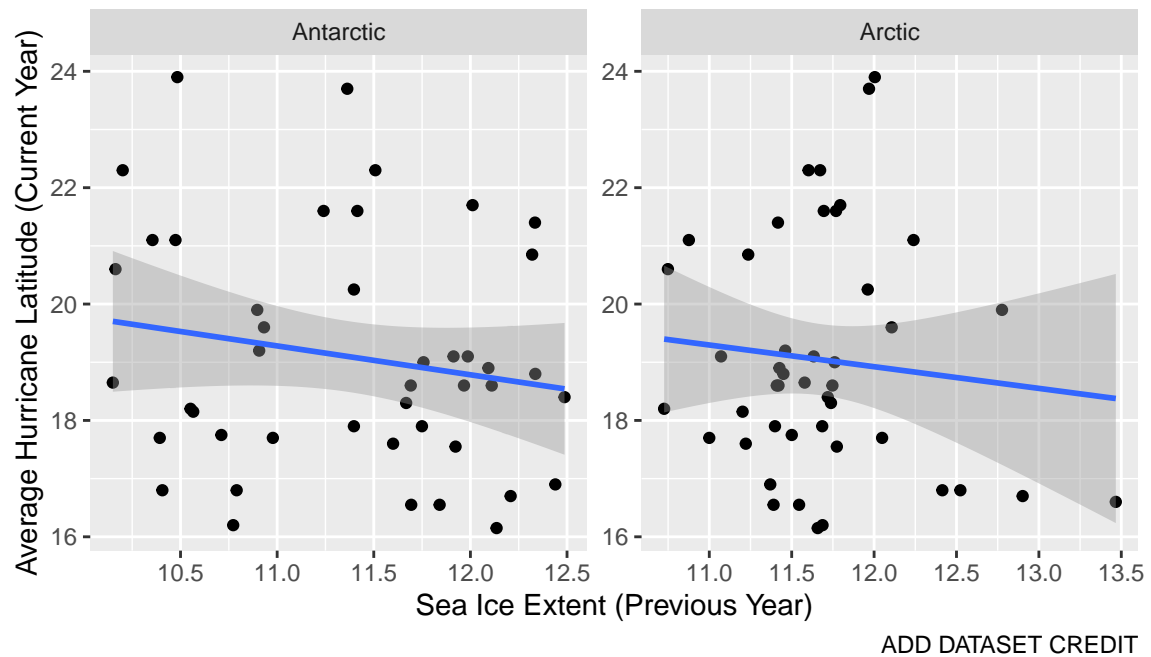[2] The relevance to the insight is . . . . From <>, published on <>, url: <>

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