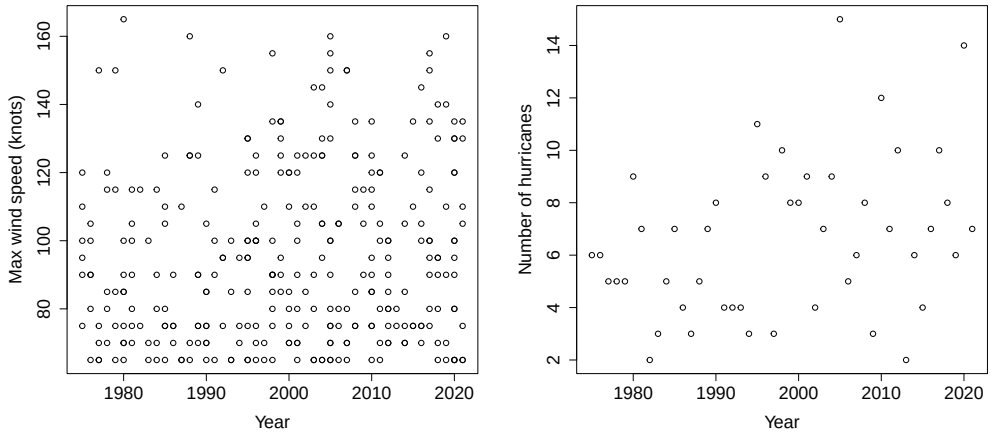
ST440/540 – Exam 1 - Due February 17, 2025

THIS IS AN EXAM - DO NOT DISCUSS THE PROBLEM WITH ANYONE (INCLUDING OTHER STUDENTS OR THE TA)! You can use the books, notes and reference manuals, but not generative AI. If you have questions, please email me.

The dataset for this exam is downloaded from the `storms` dataset in the `dplyr` package (code on the last page). The original data includes the hourly wind speed for every storm in the Atlantic. We consider only data from 1975-2021 and only storms that become hurricanes (i.e., that achieve wind speed at least 65 knots). For each storm, we compute its year (`year_storm`) and maximum wind speed (`mxspd`); these data are plotted in the left panel below. We will also consider the number of hurricanes (`cnt`) each year (`year`), plotted on the right below.



Your objective is to test whether the distribution of the maximum wind speed and/or the number of storms are changing over time. Split the years into Period 1 from 1975-1999 and Period 2 from 2000-2021 (since we haven't covered regression yet). Assume the observations are independent and identically distributed within each period, and test whether the distribution differs by period.

For all questions below, repeat the analysis separately for maximum wind speed and the number of storms (present results for wind speed on one page and the number of storms on a second page).

1. Pick parametric family of distributions for the likelihood function and confirm graphically that it fits the data reasonably well. You can consider transformations of the outcome (centering and/or scaling, log transformations, etc.), and you can treat wind speed as continuous.

2. Select conjugate uninformative prior distributions for all parameters.

3. Summarize the posterior distribution including a test to determine if the distribution differs between Periods 1 and 2, and a point estimate with standard error of the magnitude of the difference between periods. Give a Layman's terms interpreation of the point estimate.

4. Determine if the results are sensitive to the prior distribution.

5. Include a 25-50 word statement in bold that summarizes your analysis and results at the level that could appear in the *Technician*.

Your paper should be written as a professional document with full sentences, clearly labeled figures and tables and few spelling/grammatical errors. Include enough detail that the results could be replicated without looking through your code. Organize your report with subsections corresponding to the questions above. Summarize your analysis in a PDF document that is **no more than two pages long** (12 font, single space, standard margins, figues and tables count as part of the two pages); you will be penalized 10 points for each additional page. Append your code to the end of this document and submit a single document. **In-class students should turn in the exam in class on Monday, Feb 17. Online students should submit the exam on moodle.**

HAVE FUN!

Loading the data

```
library(dplyr)

# Load the data
data(storms)
storms <- storms[!is.na(storms$category),] # Only analyze hurricanes
year   <- storms$year
name   <- paste0(storms$name,"_",storms$year)
wind   <- storms$wind
uni    <- unique(name)

# Compute the maximum wind speed by storm

n          <- length(uni)
year_storm <- rep(0,n)
mxspd      <- rep(0,n)
for(i in 1:length(uni)){
      year_storm[i] <- min(year[name==uni[i]])
      mxspd[i]      <- max(wind[name==uni[i]])
}

# Compute the number of storms per year

year <- as.numeric(names(table(year_storm)))
cnt  <- as.vector(table(year_storm))

plot(year_storm,mxspd,xlab="Year",ylab="Max wind speed (knots)")
plot(year,cnt,xlab="Year",ylab="Number of hurricanes")
```