# Analyzing NHC Tropical Storm Data

Theresia Bernadette Phoa

2023-09-14

In this script, we want to analyze the National Hurricane Center (NHC) storm tracks data covering the Atlantic Basin for the years 1975 to 2021, except for the last storm in 2021 (Wanda). This dataset has been made available to us through the dplyr package.

**Questions to Address:**
1. Do Higher Wind Speeds Correspond to Wider Tropical Storm Diameter?
2. Which Months Typically Witness More Intense Storms?
3. Spatial distributions of the storms before and after year 2000

```
## Load required packages
library(tidyverse)  # For data manipulation and plotting
library(dplyr)      # For data manipulation
library(sf)         # For spatial data handling
library(OpenStreetMap)  # For working with OpenStreetMap data
library(gridExtra)  # For arranging plots

# Load the 'storms' data from dplyr package
storms
```

## Exploring the Data

```
# Check for the latest data available
tail(storms)
```

```
## # A tibble: 6 x 13
##   name   year month   day  hour   lat  long status       category  wind pressure
##   <chr> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <fct>           <dbl> <int>    <int>
## 1 Wanda  2021    11     6    18  37.1 -38   tropical st~       NA    35     1002
## 2 Wanda  2021    11     7     0  37.4 -37.4 tropical st~       NA    35     1003
## 3 Wanda  2021    11     7     6  38.1 -36.4 tropical st~       NA    35     1004
## 4 Wanda  2021    11     7    12  39.2 -34.9 other low          NA    35     1006
## 5 Wanda  2021    11     7    18  40.9 -32.8 other low          NA    40     1006
## 6 Wanda  2021    11     8     0  43.2 -29.7 other low          NA    40     1006
## # i 2 more variables: tropicalstorm_force_diameter <int>,
## #   hurricane_force_diameter <int>
```

```
# Check for recorded years and months in the data
unique(pull(storms, year))
```

```
##  [1] 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989
## [16] 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004
## [31] 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
## [46] 2020 2021
```

```
unique(pull(storms, month))
```

```
##  [1]  6  7  8  9 10 11  5 12  4  1
```

Although there are storms in every year from 1975-2021, some months may have no recorded storms

```
sort(unique(pull(storms, month)))
```

```
##  [1]  1  4  5  6  7  8  9 10 11 12
```

```
# Check recorded days in the data
sort(unique(pull(storms, day)))
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
## [26] 26 27 28 29 30 31
```

This step below might be a little bit confusing, because from tail() output we saw 2 columns for category. The 'status' column is for the storm classification (tropical wave - Hurricane ), and 'category' column is category of the hurricane based on saffir-simpson scale (1, 2, 3, 4, 5)

```
# Check what are the category of storms available.
unique(pull(storms, status))
```

```
## [1] tropical depression    tropical storm         extratropical
## [4] hurricane              subtropical storm      subtropical depression
## [7] disturbance            other low              tropical wave
## 9 Levels: disturbance extratropical hurricane ... tropical wave
```

```
# Generate a summary of the data, here we can scan through the data for any NAs
summary(storms)
```

```
##      name                year          month             day
##  Length:19066       Min.   :1975   Min.   : 1.000   Min.   : 1.00
##  Class :character   1st Qu.:1993   1st Qu.: 8.000   1st Qu.: 8.00
##  Mode  :character   Median :2004   Median : 9.000   Median :16.00
##                     Mean   :2002   Mean   : 8.699   Mean   :15.78
##                     3rd Qu.:2012   3rd Qu.: 9.000   3rd Qu.:24.00
##                     Max.   :2021   Max.   :12.000   Max.   :31.00
##
##       hour             lat             long                         status
##  Min.   : 0.000   Min.   : 7.00   Min.   :-109.30   tropical storm     :6684
##  1st Qu.: 5.000   1st Qu.:18.40   1st Qu.: -78.70   hurricane          :4684
##  Median :12.000   Median :26.60   Median : -62.25   tropical depression:3525
##  Mean   : 9.094   Mean   :26.99   Mean   : -61.52   extratropical      :2068
##  3rd Qu.:18.000   3rd Qu.:33.70   3rd Qu.: -45.60   other low          :1405
```

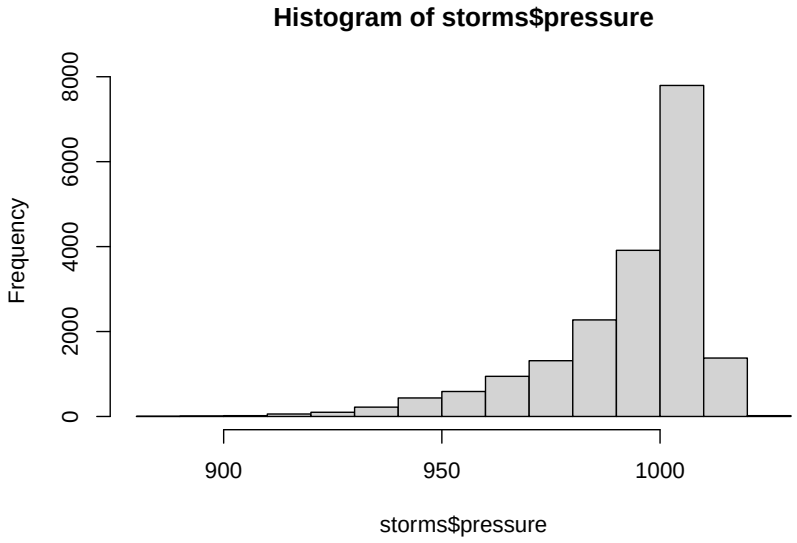
```
##  Max.   :23.000   Max.   :70.70   Max.   :  13.50   subtropical storm  : 292
##                                                     (Other)            : 408
##     category          wind           pressure      tropicalstorm_force_diameter
##  Min.   :1.000   Min.   : 10.00   Min.   : 882.0   Min.   :   0.0
##  1st Qu.:1.000   1st Qu.: 30.00   1st Qu.: 987.0   1st Qu.:   0.0
##  Median :1.000   Median : 45.00   Median :1000.0   Median : 110.0
##  Mean   :1.898   Mean   : 50.02   Mean   : 993.6   Mean   : 146.3
##  3rd Qu.:3.000   3rd Qu.: 65.00   3rd Qu.:1007.0   3rd Qu.: 220.0
##  Max.   :5.000   Max.   :165.00   Max.   :1024.0   Max.   :1440.0
##  NA's   :14382                                     NA's   :9512
##  hurricane_force_diameter
##  Min.   :  0.00
##  1st Qu.:  0.00
##  Median :  0.00
##  Mean   : 14.81
##  3rd Qu.:  0.00
##  Max.   :300.00
##  NA's   :9512
```

Tropical storm & hurricane force wind diameter data are not available until 2004.

```
# Create a copy of the 'storms' data in a new dataframe to work with,
# so we don't have to download it everytime we want to use the script
# (e.g., data cleaning, data manipulation, etc.)
storms <- storms
table(storms$status)
```

```
##
##            disturbance          extratropical              hurricane
##                    146                   2068                   4684
##              other low subtropical depression      subtropical storm
##                   1405                    151                    292
##    tropical depression         tropical storm          tropical wave
##                   3525                   6684                    111
```

```
# Create a histogram of 'pressure' values
hist(storms$pressure)
```



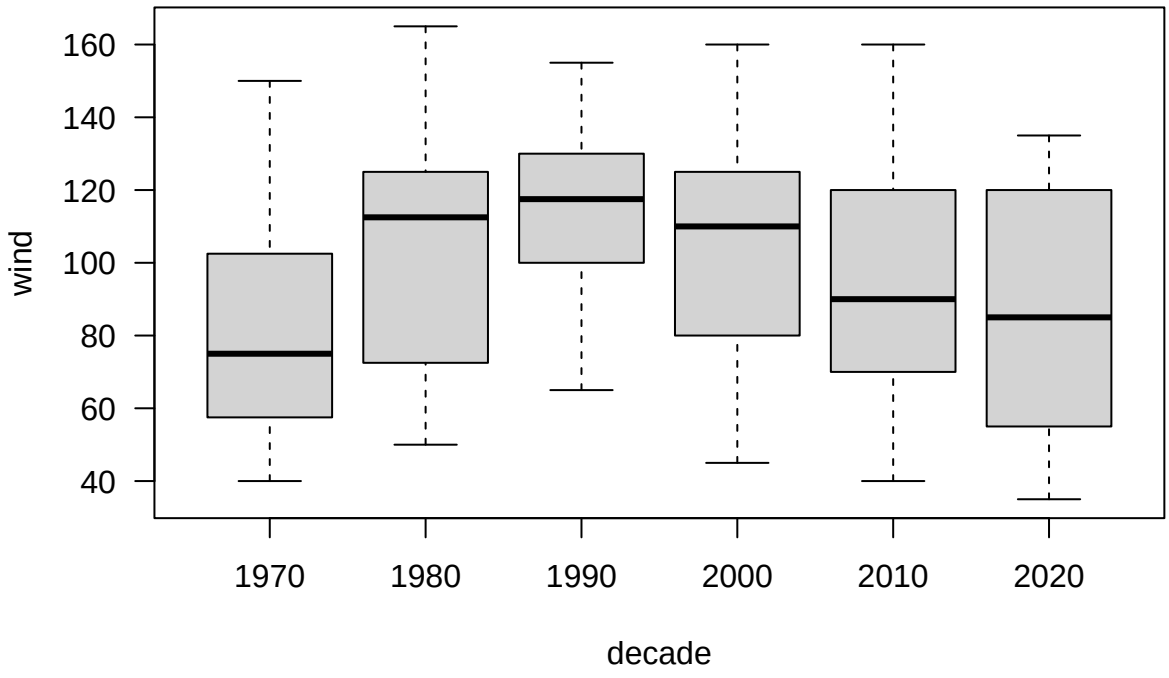

Now, we want to group the data by storm name, filter to get minimum pressure and maximum wind for each storm. This data is originally a point observation data, we can use it for track visualization. However, to analyze the storm events statistics won't be effective if we don't summarize the data.

```
storm_summary <- storms %>%
  group_by(name) %>%
  filter(pressure == min(pressure)) %>%
  filter(status %in% c("tropical storm", "hurricane")) %>%
  filter(wind == max(wind)) %>%
  mutate(date = as.POSIXct(paste(year, month, day, hour, sep = "-"),
                           format = "%Y-%m-%d-%H")) %>%
  filter(date == max(date))
```

We also filter the data only for recorded tropical storms (>34 kts) and hurricanes (>64 kts)

```
# Create 'decade' column to group storms by decade
storm_summary <- storm_summary %>%
  mutate(decade = 10 * (year %/% 10))
```

```
# Create a boxplot of wind speeds by decade
boxplot(wind ~ decade, data = storm_summary, las = 1)
```

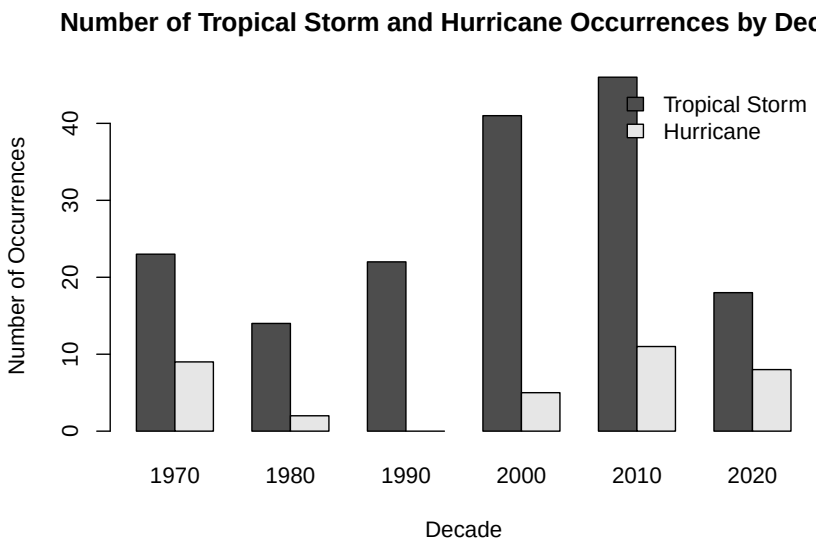



```
# Create a summary dataframe for number of tropical storms and hurricanes by decade
summary_decade <- storm_summary %>%
  group_by(decade, status) %>%
  summarize(count = n())
```

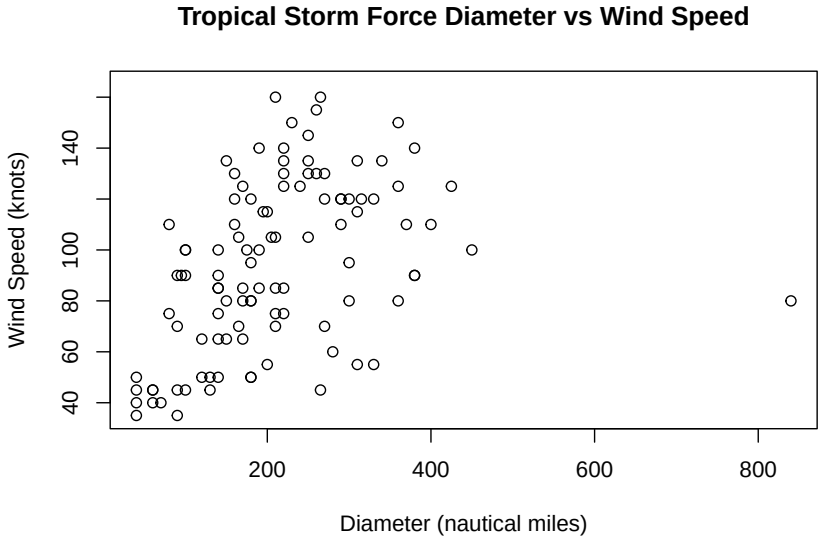
```
## `summarise()` has grouped output by 'decade'. You can override using the
## `.groups` argument.
```

```
# Pivot the summary data to wide format
summary_wide <- summary_decade %>%
  pivot_wider(names_from = status, values_from = count, values_fill = 0)

# Create a matrix for the barplot
decade_matrix <- t(as.matrix(summary_wide[, -1]))
```

```
# Create a barplot
barplot(decade_matrix, beside = TRUE,
        names.arg = unique(summary_wide$decade),
        xlab = "Decade",
        ylab = "Number of Occurrences",
        main = "Number of Tropical Storm and Hurricane Occurrences by Decade",
        legend.text = c("Tropical Storm", "Hurricane"),
        args.legend = list(x = "topright", bty = "n"))
```




```
# Create a scatterplot of tropical storm force diameter vs wind speed
plot(storm_summary$tropicalstorm_force_diameter,
     storm_summary$wind,
     xlab = "Diameter (nautical miles)",
     ylab = "Wind Speed (knots)",
     main = "Tropical Storm Force Diameter vs Wind Speed")
```




```
# Calculate the correlation coefficient between wind speed and tropical
# storm force diameter.
correlation_coefficient <- cor(storm_summary$wind,
                               storm_summary$tropicalstorm_force_diameter,
                               use = "complete.obs")
cat("Correlation Coefficient:", correlation_coefficient, "\n")
```

```
## Correlation Coefficient: 0.4079453
```

```
# Find the storm with the largest tropical storm force diameter
largest_diameter_name <- storm_summary$name[which.max(storm_summary$tropicalstorm_force_diameter)]
largest_diameter_name
```

```
## [1] "Sandy"
```

```
# Create a summary of hurricane counts by month
month_hurricane <- storm_summary %>%
  filter(category %in% c(1:6)) %>%
  group_by(month) %>%
  summarize(hurricane_count = n())

# Find the peak hurricane season
peak_season_hurricane <- month_hurricane %>%
  arrange(desc(hurricane_count)) %>%
  head(1)

peak_season_hurricane
```

```
## # A tibble: 1 x 2
##   month hurricane_count
##   <dbl>           <int>
## 1     9              70
```

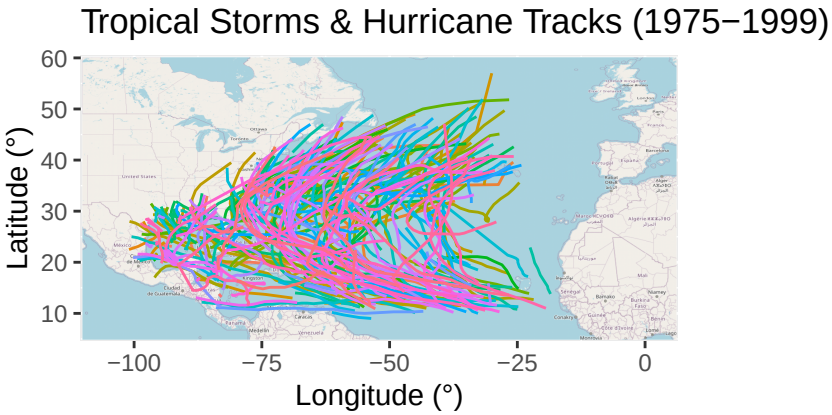
```
# Calculate correlation coefficients between wind speed > 64 knots
# and each month
correlation_coefficients <- sapply(1:12, function(i) {
  cor(storm_summary$wind[storm_summary$wind > 64],
      storm_summary$month[storm_summary$wind > 64] == i,
      method = "spearman")
})

# Create a dataframe to store the correlation coefficients and
# corresponding months
correlation_df <- data.frame(
  Month = 1:12,
  Correlation_Coefficient = correlation_coefficients
)

# Print the correlation coefficients
correlation_df
```

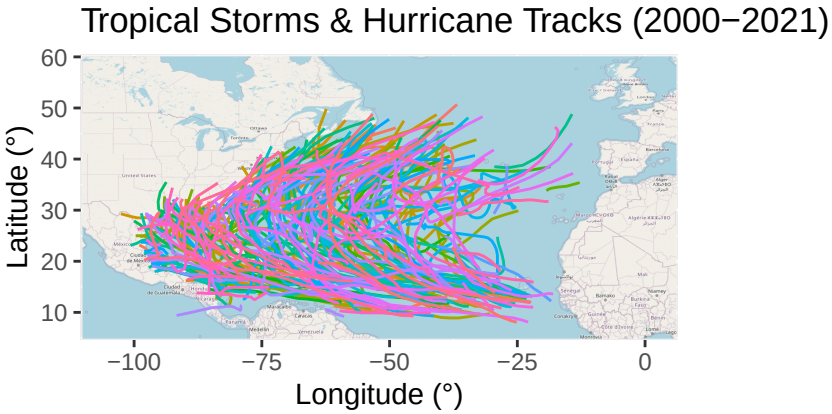
```
##    Month Correlation_Coefficient
## 1      1                      NA
## 2      2                      NA
## 3      3                      NA
## 4      4                      NA
## 5      5                      NA
## 6      6                      NA
## 7      7            -0.136447985
```

```
## 8      8             0.071617789
## 9      9             0.161365343
## 10    10            -0.173892113
## 11    11             0.003966381
## 12    12                      NA
```

**Spatial distributions of the storms before and after year 2000**

```r
# Subset the 'storms' data for years before 2000 and after 2000
storms_before_2000 <- storms %>%
  filter(year < 2000) %>%
  filter(status %in% c("tropical storm", "hurricane"))

storms_after_2000 <- storms %>%
  filter(year >= 2000) %>%
  filter(status %in% c("tropical storm", "hurricane"))

# Define latitudes and longitudes for the map
lat1 <- 5; lat2 <- 60; lon1 <- -110; lon2 <- 6

# Create a base map using OpenStreetMap
base_map <- openmap(c(lat2, lon1), c(lat1, lon2), zoom = 4, type = "osm",
                    mergeTiles = TRUE)
base_map2 <- openproj(base_map)

# Create a plot for hurricane tracks on top of the base map
plot_before_2000 <- OpenStreetMap::autoplot.OpenStreetMap(base_map2) +
  geom_path(data = storms_before_2000, aes(x = long, y = lat, color = paste(year, name)),
            show.legend = FALSE) +
  xlab("Longitude (°)") + ylab("Latitude (°)") +
  labs(title = "Tropical Storms & Hurricane Tracks (1975-1999)")

plot_after_2000 <- OpenStreetMap::autoplot.OpenStreetMap(base_map2) +
  geom_path(data = storms_after_2000, aes(x = long, y = lat, color = paste(year, name)),
            show.legend = FALSE) + xlab("Longitude (°)") + ylab("Latitude (°)") +
  labs(title = "Tropical Storms & Hurricane Tracks (2000-2021)")

# Arrange the plots side by side
grid.arrange(plot_before_2000, plot_after_2000, nrow = 2)
```




**Other statistical analysis (likelihood of trop. storm events per extended period)**

```r
# Calculate Poisson distribution
# Filter the data frame to include only category 4 or 5 hurricanes
storm_summary_cat4 <- storm_summary %>%
  filter(category == '4')

storm_summary_cat5 <- storm_summary %>%
  filter(category == '5')

# Calculate the total number of category 4 and 5 hurricanes over the decade
total_category4 <- nrow(storm_summary_cat4)
total_category5 <- nrow(storm_summary_cat5)

# Calculate the average number of category 4 and 5 hurricanes per year
average_category4_per_year <- total_category4 / 46
average_category5_per_year <- total_category5 / 46

# Create a sequence of numbers for plotting the Poisson distribution
x = 1:50

# Calculate the Poisson probabilities for category 4 and 5 hurricanes
poisson_cat4 <- dpois(x, lambda = average_category4_per_year)
poisson_cat5 <- dpois(x, lambda = average_category5_per_year)
```

```r
# Create a two-panel plot for Poisson distribution
par(mfrow = c(1, 2))
plot(x, poisson_cat4, type = "l", lwd = 3, col = "navy")
plot(x, poisson_cat5, type = "l", lwd = 3, col = "navy")
```