ProjectFinal

Sandu Galbokka Hewa

2023-12-08

The Dataset of Storms from 1975-2021

This dataset is the NOAA Atlantic hurricane database best track data.

First we look simply what we got in this dataset.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
sample_n(storms, 5)
```

```
## # A tibble: 5 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 Alberto  2012     5    21     6  30.4 -79.8 tropical ~       NA    35     1007
## 2 Alex     1998     7    28    12  12.9 -33.3 tropical ~       NA    30     1007
## 3 Fran     1996     9     4     6  26.4 -73.9 hurricane         3   100      953
## 4 Chris    1994     8    18     6  14.5 -46.8 hurricane         1    65      987
## 5 David    1979     9     4     0  28   -80.5 hurricane         2    85      971
## # i 2 more variables: tropicalstorm_force_diameter <int>,
## #   hurricane_force_diameter <int>
```

```
str(storms)
```

```
## tibble [19,066 x 13] (S3: tbl_df/tbl/data.frame)
##  $ name                        : chr [1:19066] "Amy" "Amy" "Amy" "Amy" ...
##  $ year                        : num [1:19066] 1975 1975 1975 1975 1975 ...
```

```
##  $ month                      : num [1:19066] 6 6 6 6 6 6 6 6 6 6 ...
##  $ day                        : int [1:19066] 27 27 27 27 28 28 28 28 29 29 ...
##  $ hour                       : num [1:19066] 0 6 12 18 0 6 12 18 0 6 ...
##  $ lat                        : num [1:19066] 27.5 28.5 29.5 30.5 31.5 32.4 33.3 34 34.4 34 ...
##  $ long                       : num [1:19066] -79 -79 -79 -79 -78.8 -78.7 -78 -77 -75.8 -74.8 ...
##  $ status                     : Factor w/ 9 levels "disturbance",..: 7 7 7 7 7 7 7 7 8 8 ...
##  $ category                   : num [1:19066] NA NA NA NA NA NA NA NA NA NA ...
##  $ wind                       : int [1:19066] 25 25 25 25 25 25 25 30 35 40 ...
##  $ pressure                   : int [1:19066] 1013 1013 1013 1013 1012 1012 1011 1006 1004 1002 ...
##  $ tropicalstorm_force_diameter: int [1:19066] NA NA NA NA NA NA NA NA NA NA ...
##  $ hurricane_force_diameter   : int [1:19066] NA NA NA NA NA NA NA NA NA NA ...
```

The data includes the positions and attributes of storms from 1975-2021. Storms from 1979 onward are measured every six hours during the lifetime of the storm. This have 19066 observations and 13 variables.

- name: Storm Name
- year,month,day: Date of report
- hour: Hour of report (in UTC)
- lat,long: Location of storm center
- status: Storm classification (Tropical Depression, Tropical Storm, or Hurricane)
- category: Saffir-Simpson hurricane category calculated from wind speed.

– NA: Not a hurricane

– 1: 64+ knots

– 2: 83+ knots

– 3: 96+ knots

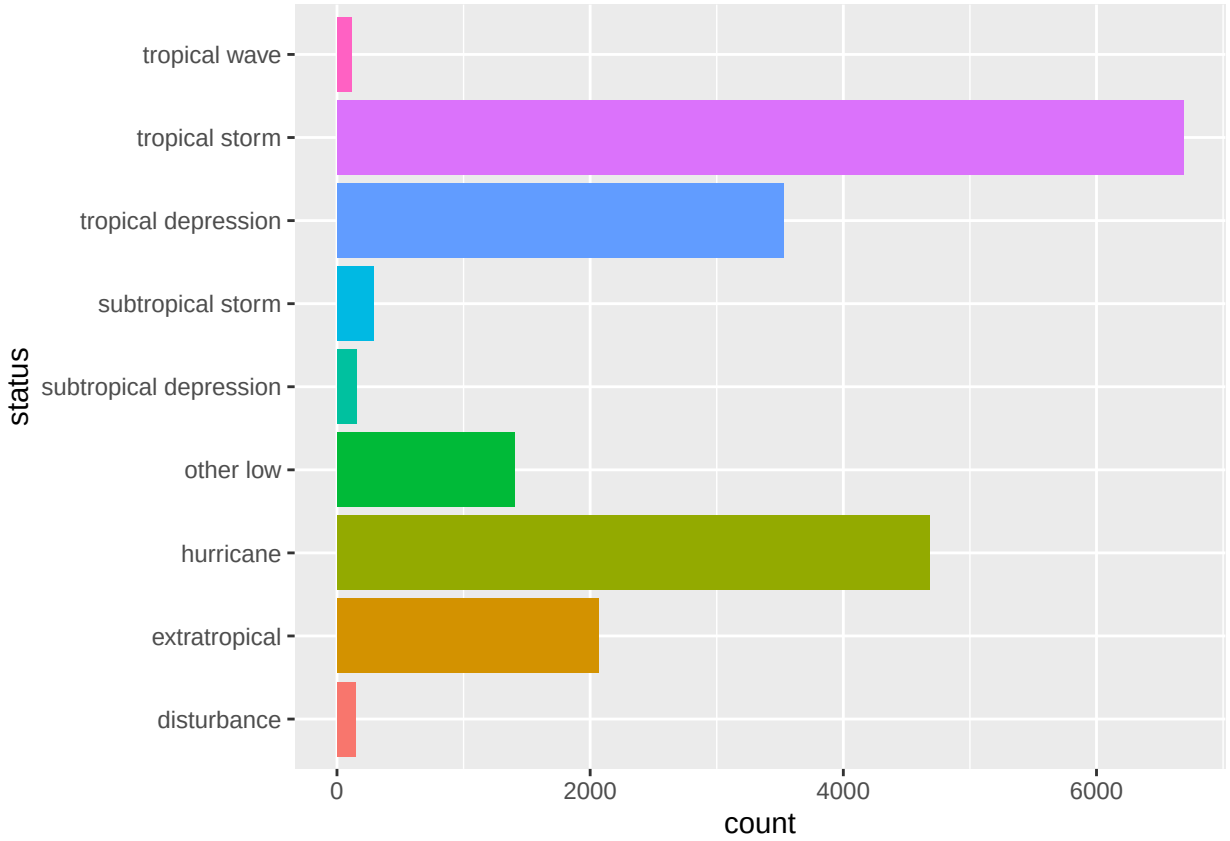
– 4: 113+ knots

– 5: 137+ knots

- wind: storm's maximum sustained wind speed (in knots)
- pressure: Air pressure at the storm's center (in millibars)
- tropicalstorm_force_diameter: Diameter (in nautical miles) of the area experiencing tropical storm strength winds (34 knots or above). Only available starting in 2004.
- hurricane_force_diameter: Diameter (in nautical miles) of the area experiencing hurricane strength winds (64 knots or above). Only available starting in 2004.

Let's see what the kind of storms this dataset has and how many storms each kind has.

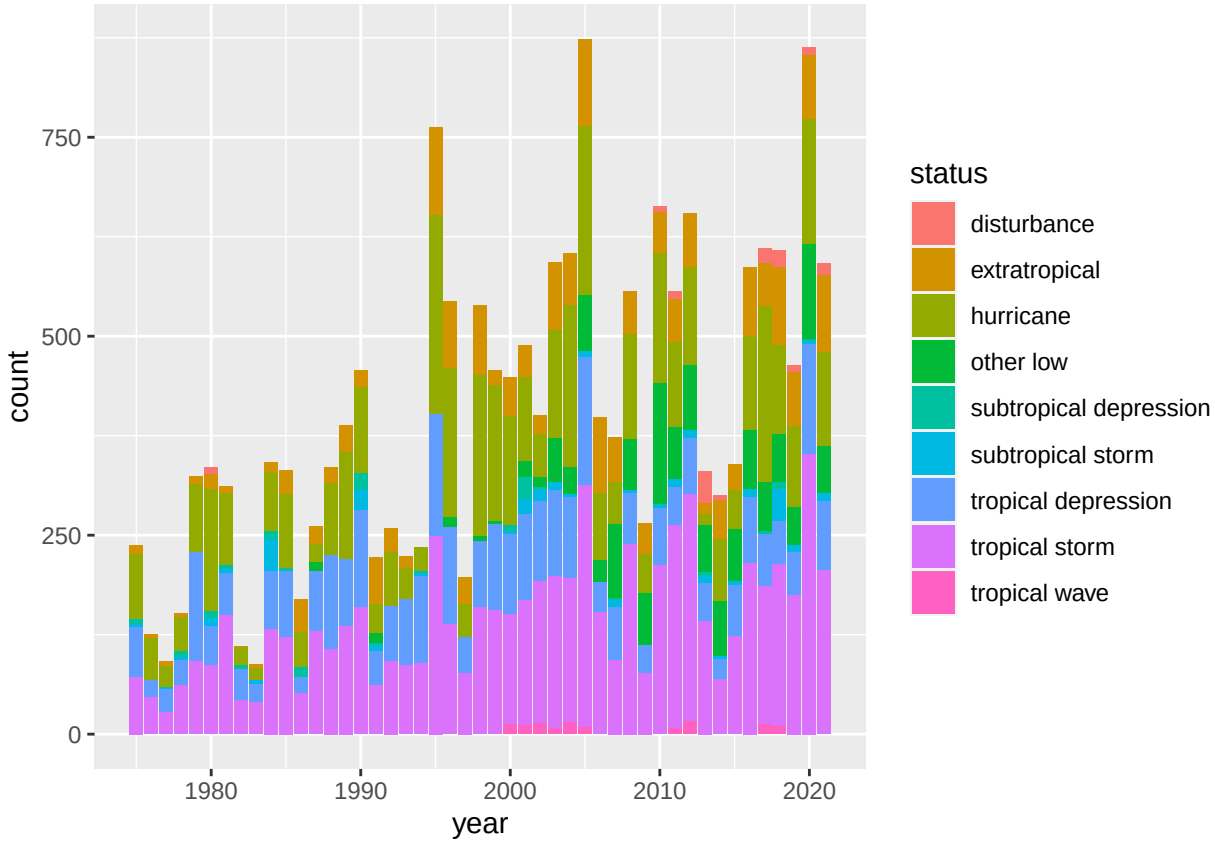
```
ggplot(data = storms, mapping = aes(x=status, fill = status)) +
  geom_bar() + guides(fill = "none") + coord_flip()
```



Above bar graph shows that tropical storm happened most followed by hurricane, while tropical wave category is the least category that happened from 1975-2021.

Next, see how many storms occurred in each year from 1975 - 2021.

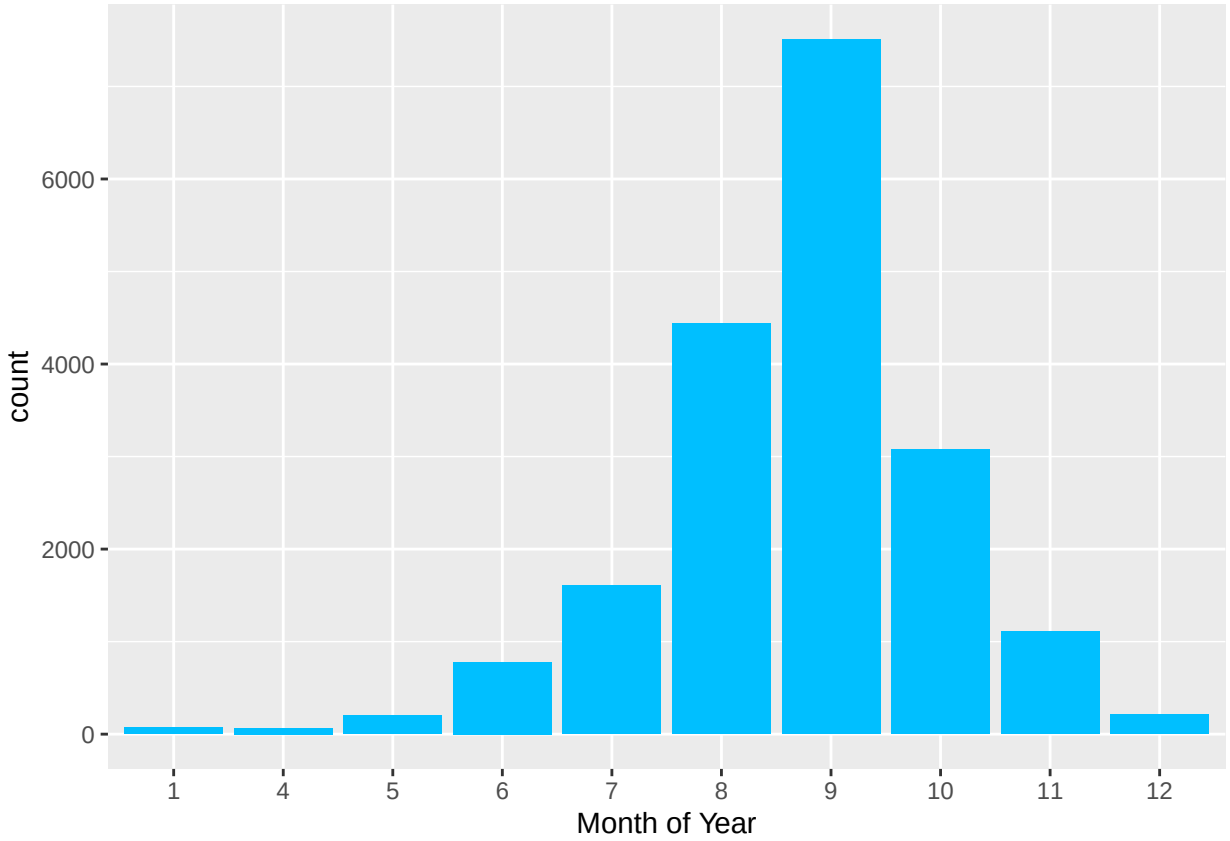
```
ggplot(data = storms, mapping = aes(x=year, fill = status)) +
  geom_bar()
```



It seems there were more storms happened after 1994 than before 1995. The year of most storms(about 875 storms) happened was 2005 followed by the year 2020(about 850 storms).

The below graph shows how occationally storms have happened throughout a year.

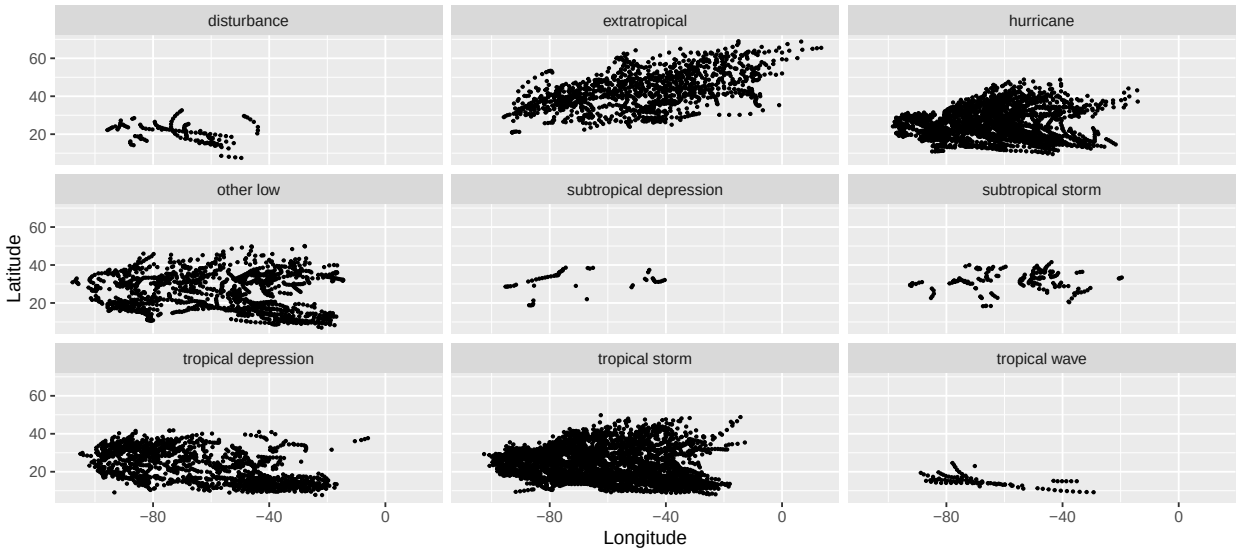
```
ggplot(data = storms, mapping= aes(x=factor(month))) +
  geom_bar(fill = "deepskyblue") + labs(x = "Month of Year")
```



Let's wrap up what we can get from above graph. - The peak activity in Atlantic storms could appear were September and August respectively. - We can roughly say as mentioned in the NOAA Atlantic hurricane database more active months for storms were June to November. - There was no Atlantic storms happened in the months of February and March. We can assume that typically in wintry months of February and March would likely be expecting a snowstorm rather than a tropical storm as the reason.

Now, will find out roughly which area of latitude and longitude each type of storms happened from 2000-2021.

```
storms %>%
  filter(year > 2000) %>%
  ggplot( mapping = aes(x = long, y = lat)) +
  geom_point(aes(group = status), size= 0.5) + facet_wrap( vars(status)) +
  labs(x= "Longitude",
       y = "Latitude")
```



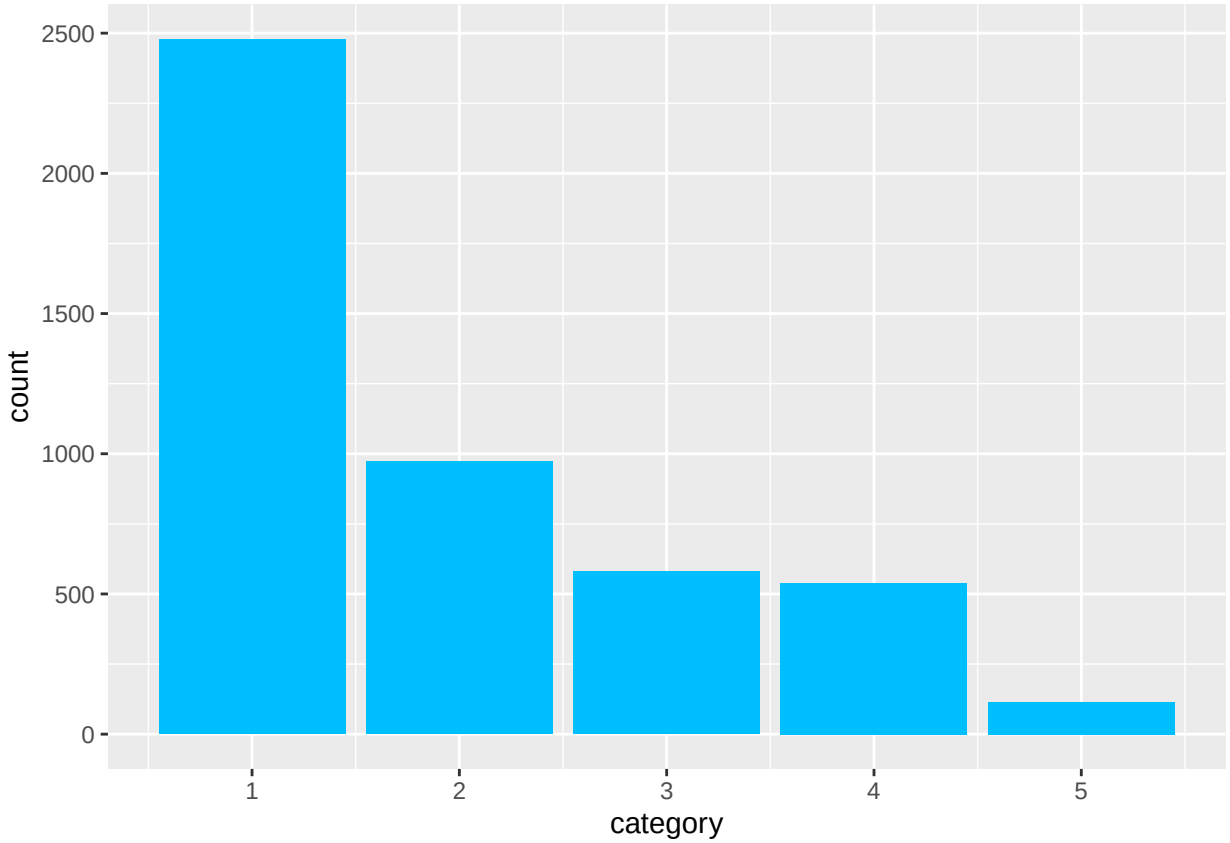
From above graph we can notice that latitude between 0-50 and longitude between -100 and - 20 was the location where mass amount of storms especially hurricane, other low, tropical depression and tropical storm had begun.

But when we focus on extratropical storms, it seems that those storms were mostly at latitude between 30 and 70 and longitude -100 and 10. They have categorized these storms by the place where storms began. For example, the extratroplical storms tells that the descriptor extratropical signifies that this type of cyclone generally occurs outside the tropics and in the middle latitudes of the Earth between 30 and 60 latitude.

Let's talk about Hurricane storms..

There is a scale named Saffir-Simpson Hurricane Wind Scale which is a 1 to 5 rating calculated from hurricane's wind speed. Below bar graph will show how many hurricanes were in each category.

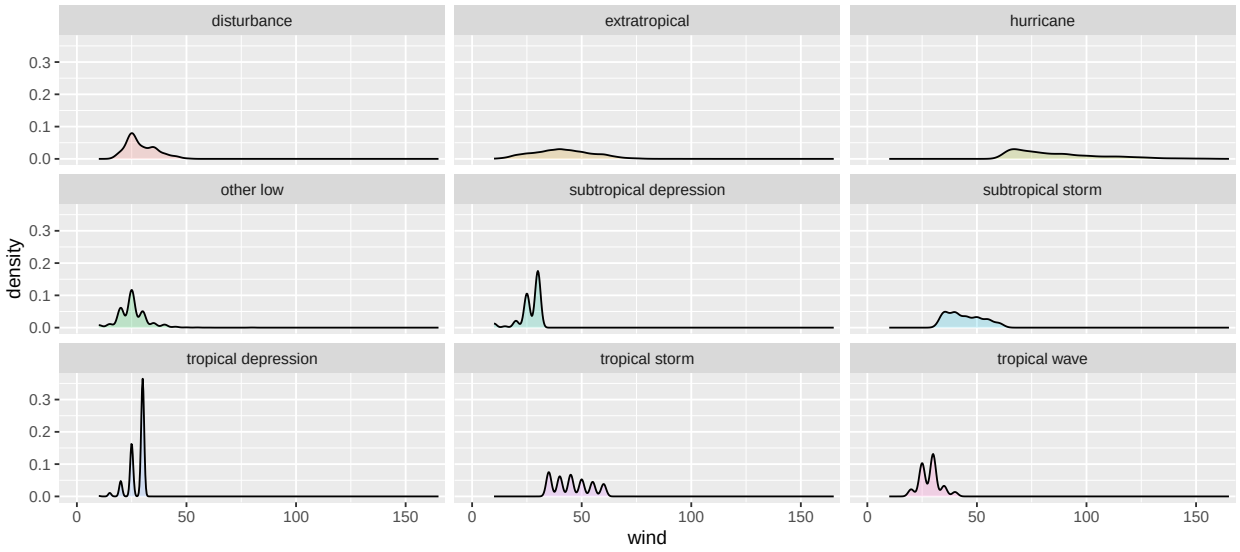
```
storms %>%
  filter(!is.na("NA")) %>%
  ggplot(mapping = aes(x = category)) +
  geom_bar(fill = "deepskyblue")
```



Almost 2500, which is the highest number, hurricanes were category 1 hurricanes, while about 250 hurricanes were category 5 hurricanes.

Below 9 graphs show how the storm's maximum sustained wind speed(in knots) varies with the type of storm.

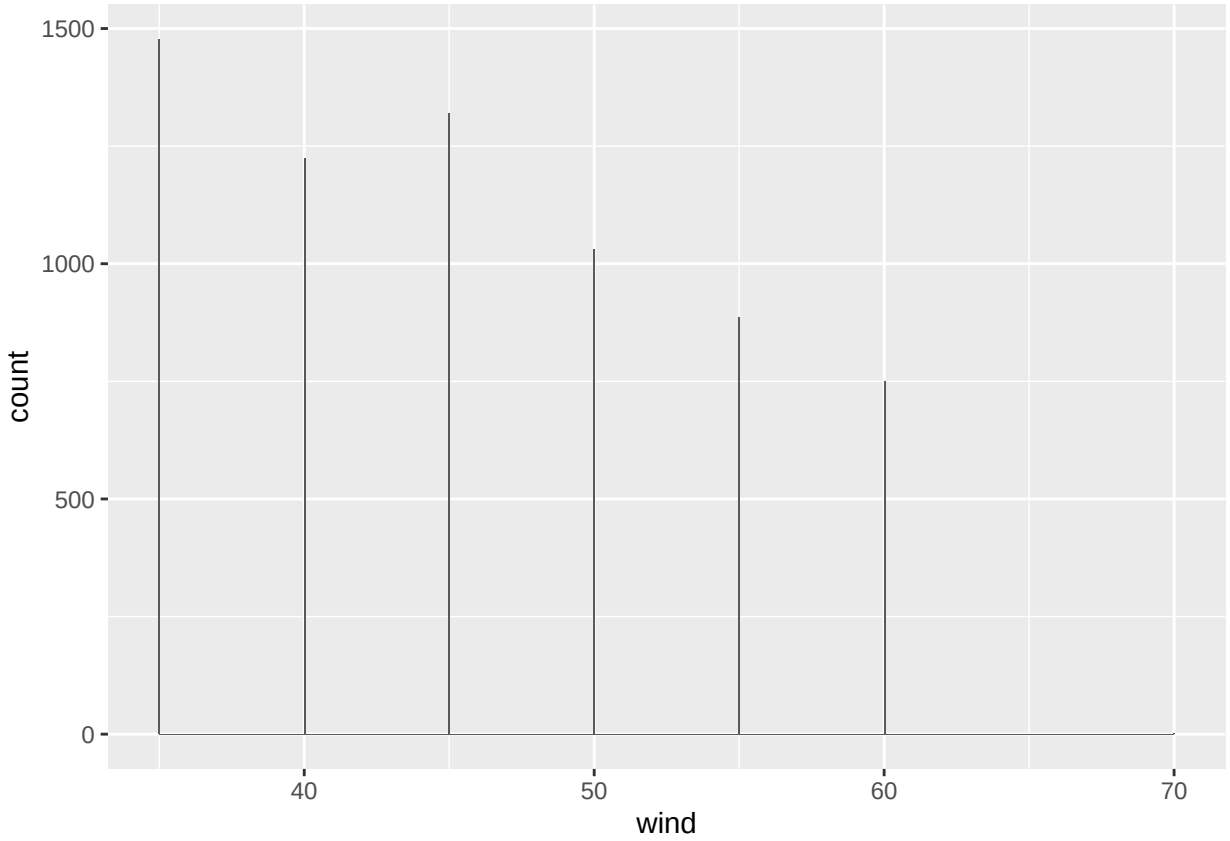
```
storms %>%
  ggplot(mapping = aes(x = wind, fill = status)) +
  geom_density(alpha = 0.2) + guides(fill = "none")+ coord_cartesian(xlim = c(0,160)) + facet_wrap( vars
```



After observing above graphs we can see that except hurricane, - Every other types of storms had the maximum sustained wind speed lower than 75 in knots. - More especificly, disturbance, tropical wave, other low, subtropical depression, and tropical depression had the the maximum sustained wind speed lower than about 50 in knots. - Hurricanes seem to have more wind speed, and I believe that's why there is a special scale for only hurricane storms to observe its wind speed. The most prominent characteristic of tropical storms is having the highest number of storms with less wind speed which is about 30 in knots. - Also, there is a strange pattern can be seen in tropical storm.

So let's see in deep what happened in tropical storm's wind.

```
storms %>%
  filter(status == "tropical storm") %>%
  ggplot(mapping = aes(x = wind)) +
  geom_histogram(bins = 1000)
```

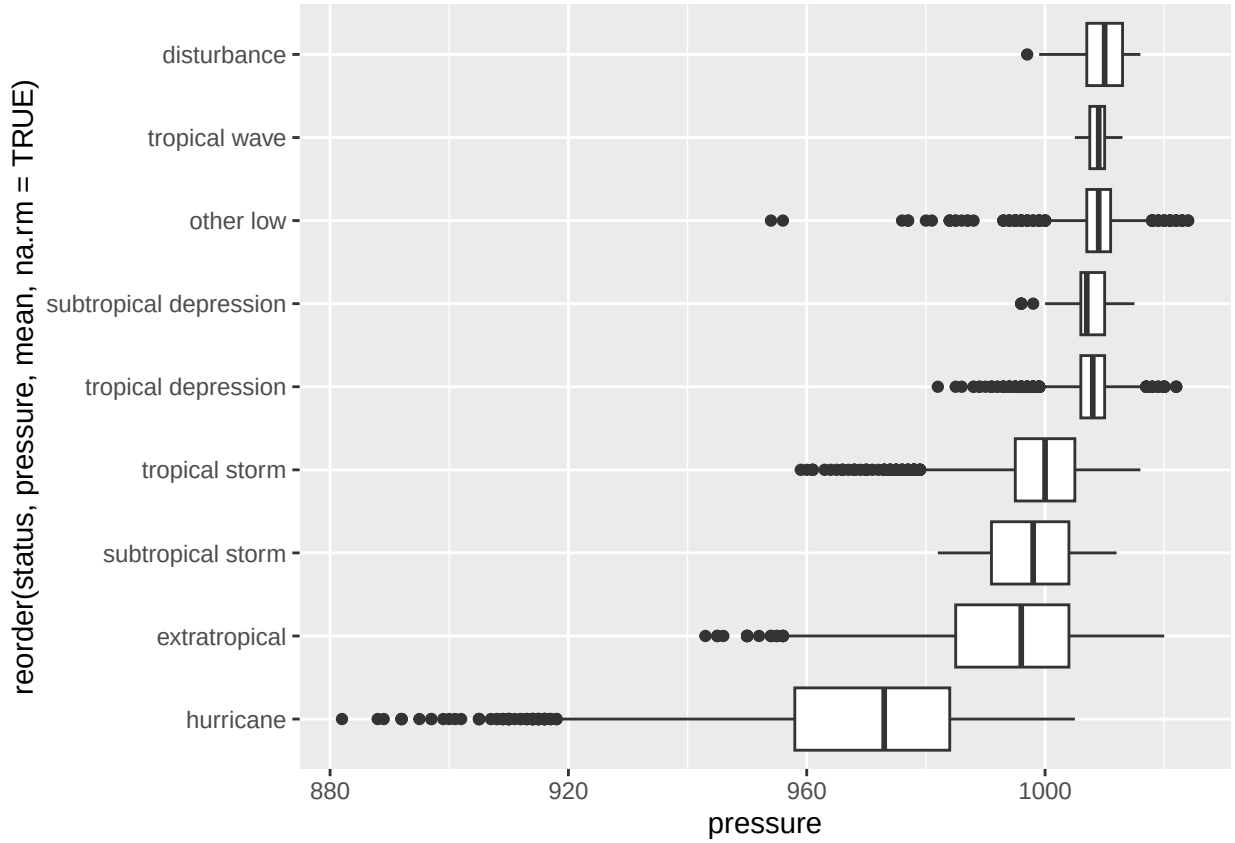


The reason for strange pattern in tropical storms is that its maximum sustained wind speeds were a five by five pattern.

Below boxplot shows how the storm's air pressure at the storm's center(in millibars) varies with the type of storm

Additionally it shows which type has the highest mean of pressure.

```
storms %>%
  ggplot(mapping = aes(x = pressure, y = reorder(status, pressure, mean, na.rm =TRUE))) +
    geom_boxplot()
```



Once looked at above boxplot we can seperate easily these 9 types in to 3 categories by looking at their meadians in presure. - 1st category - disturbance, tropical wave, other low, subtropical depression, tropical depression - 2nd category - tropical storm, subtropical storms, extratropical - 3rd category - hurricane

If we compare these three categories's pressure with the graph of its maximum sustained wind speed(in knots), we can notice, hurricane(3rd category) generally has the highest wind speed and the lowest meadian in pressure. the storm types in 1st category have the lowest wind speed and the highest meadian in pressure.

These observations bring us that there is a connection between the wind of storm and its center pressure.

Let's draw a linear regression for pressure vs wind to find this connection.

```
mod1 <- lm(pressure ~ wind, data = storms)
mod1
```

```
##
## Call:
## lm(formula = pressure ~ wind, data = storms)
##
## Coefficients:
## (Intercept)         wind
##   1027.7029      -0.6827
```

```
summary(mod1)
```

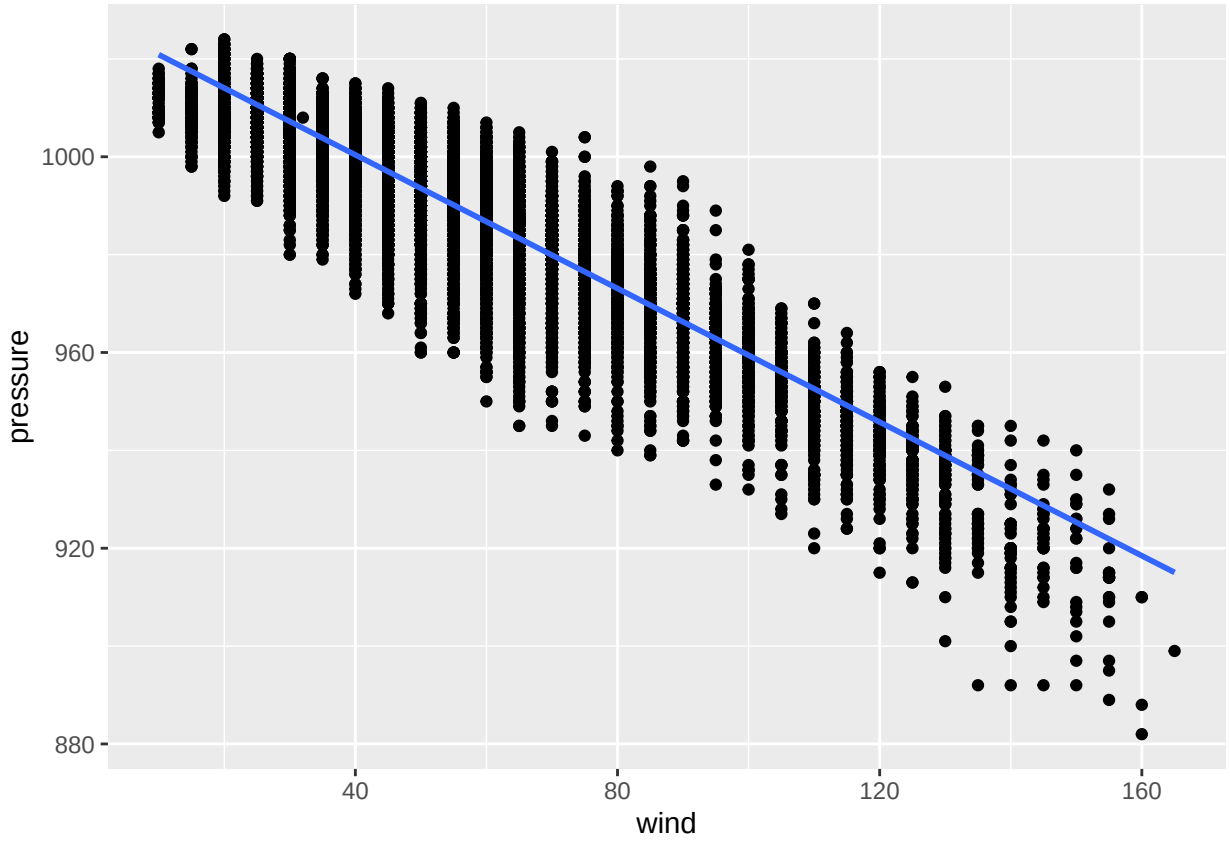
```
##
## Call:
```

```
## lm(formula = pressure ~ wind, data = storms)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.534  -3.016   0.916   3.847  28.743
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.028e+03  1.104e-01  9306.2   <2e-16 ***
## wind        -6.827e-01  1.967e-03  -347.1   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.926 on 19064 degrees of freedom
## Multiple R-squared:  0.8634, Adjusted R-squared:  0.8634
## F-statistic: 1.205e+05 on 1 and 19064 DF,  p-value: < 2.2e-16
```

See what we got. The adjusted R-squared is 0.8634. It means **86.34%** of variability of pressure is explained by the value of x which is wind. Still about 13% of another variable/variables affects for wind. Probably temperature may be the another impact for wind.

```
ggplot(mod1) +
  geom_point(mapping = aes(x = wind, y=pressure)) +
  geom_smooth(aes(x = wind, y=pressure),  method = "lm")
```

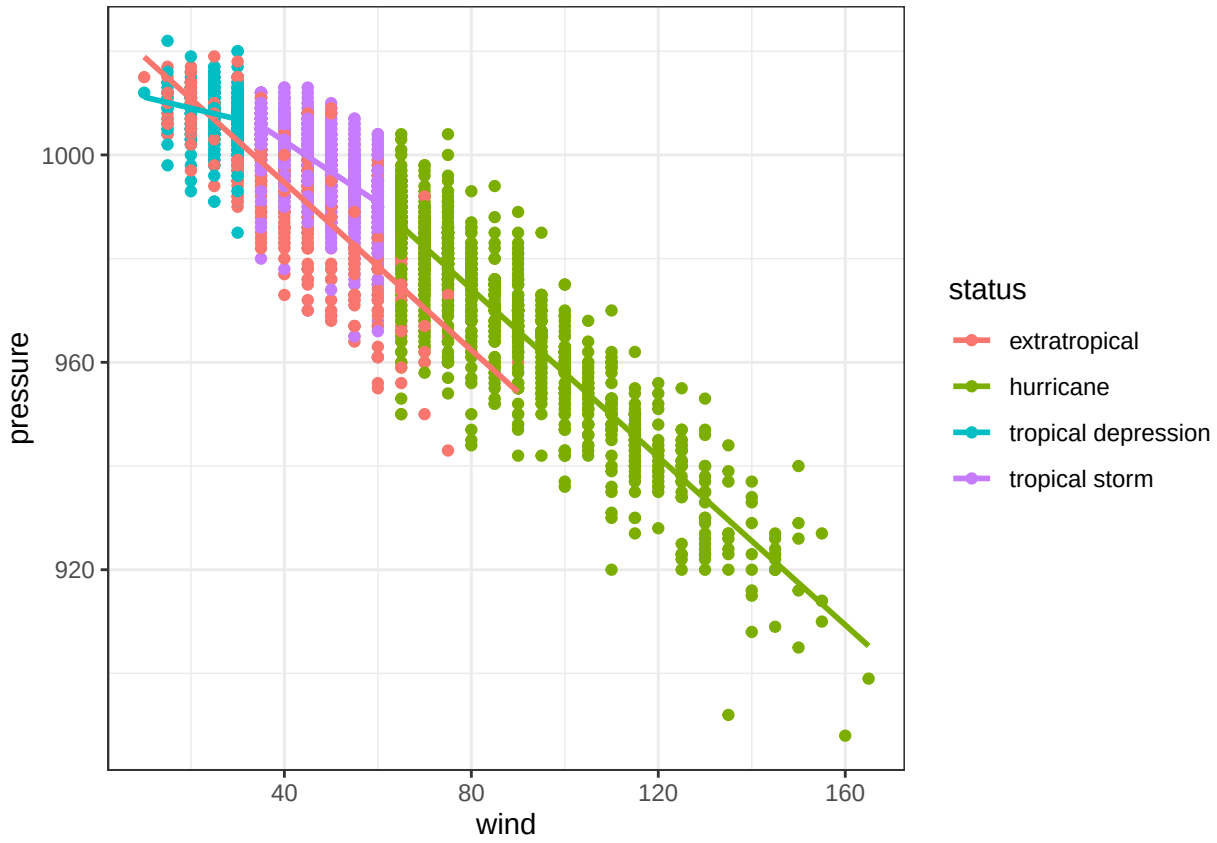
```
## `geom_smooth()` using formula = 'y ~ x'
```



How ecah type's wind varies with its pressure...

```
storms %>%
  filter(status== c("tropical storm", "tropical depression", "extratropical", "hurricane")) %>%
  ggplot(aes(x = wind, y = pressure, color = status)) +
    theme_bw() +
    geom_point() +
    geom_smooth(method = 'lm', se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



It seems hurricane and extratropical have same slop but different intercept. Also, other two types of storms which appear in the chart, have negative slops.

However, still we can conclude that if the storm's pressure is low, its wind will be high since all slopes are negative slopes.