

Final Project

Kenan

12/24/2020

Storm track dataset

As the planet temperature has been increasing due to human-made global warming it has created conditions that increase the chances of extreme weather. This report is will investigate the claim of increasing storm rates.

Description This data is a subset of the NOAA Atlantic hurricane database best track data, <https://www.nhc.noaa.gov/data/#hurdat>. The data includes the positions and attributes of 198 tropical storms, measured every six hours during the lifetime of a storm.

- Format - A tibble with 10,010 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 storm category (estimated from wind speed. -1 = Tropical Depression, 0 = Tropical Storm)
- wind - storm's maximum sustained wind speed (in knots)
- pressure - Air pressure at the storm's center (in millibars)
- ts_diameter - Diameter of the area experiencing tropical storm strength winds (34 knots or above)
- hu_diameter - Diameter of the area experiencing hurricane strength winds (64 knots or above) "

Read in the dataset and remove unnecessary columns Begin by reading the data and performing simple data cleaning operations

```
df <- read.csv('https://vincentarelbundock.github.io/Rdatasets/csv/dplyr/storms.csv')
df$date <- paste(df$year, df$month, df$day, sep='-')
df <- df[, !(names(df) %in% c('month', 'day', 'hour', 'X'))]
df <- df %>% replace(is.na(.), 0)
df$total_diameter <- df$ts_diameter + df$hu_diameter
```

Global statistics

```
summary(df)
```

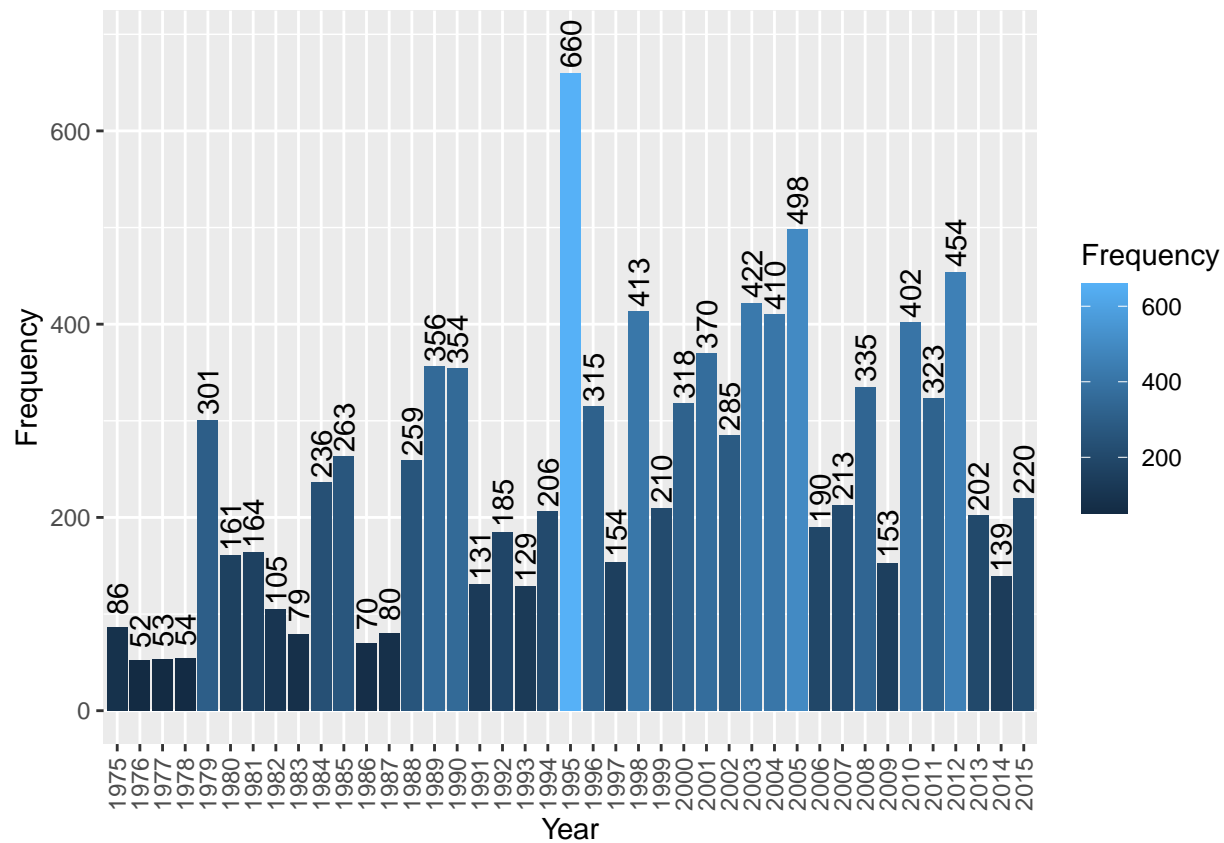
##	name	year	lat	long
##	Emily	: 207 Min. :1975	Min. : 7.20	Min. : -109.30
##	Bonnie	: 185 1st Qu.:1990	1st Qu.:17.50	1st Qu.: -80.70

```
## Claudette: 180   Median :1999   Median :24.40   Median : -64.50
## Felix   : 178   Mean    :1998   Mean    :24.76   Mean    : -64.23
## Alberto : 170   3rd Qu.:2006   3rd Qu.:31.30   3rd Qu.: -48.60
## Danielle: 157   Max.    :2015   Max.    :51.90   Max.    :  -6.00
## (Other) :8933

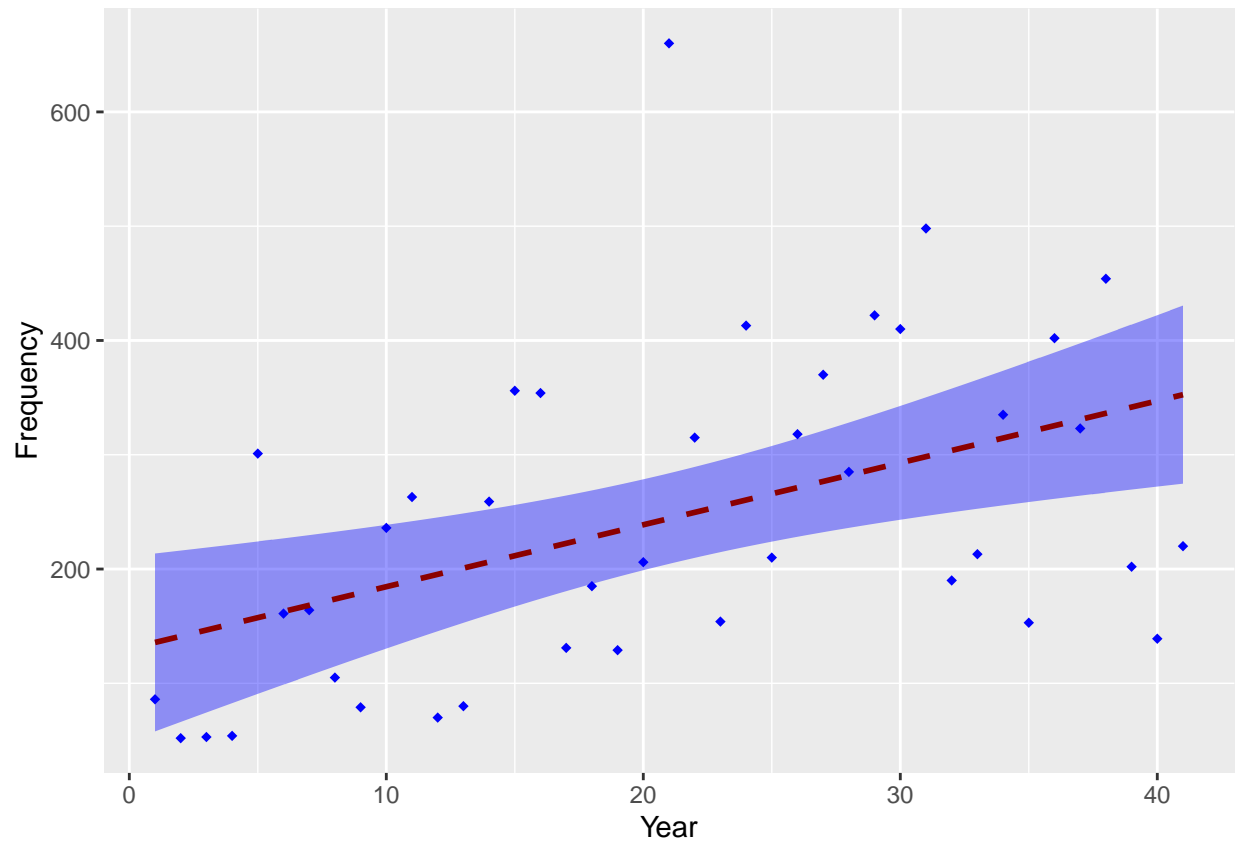
##           status      category      wind      pressure
## hurricane      :3091   Min.     :-1.0000   Min.     : 10.00   Min.     : 882.0
## tropical depression:2545 1st Qu.: -1.0000   1st Qu.: 30.00   1st Qu.: 985.0
## tropical storm   :4374   Median : 0.0000   Median : 45.00   Median : 999.0
##                  Mean    : 0.3214   Mean    : 53.49   Mean    : 992.1
##                  3rd Qu.: 1.0000   3rd Qu.: 65.00   3rd Qu.:1006.0
##                  Max.    : 5.0000   Max.    :160.00   Max.    :1022.0
##
##   ts_diameter   hu_diameter      date      total_diameter
## Min.   : 0.00   Min.   : 0.000   Length:10010   Min.   : 0.00
## 1st Qu.: 0.00   1st Qu.: 0.000   Class :character   1st Qu.: 0.00
## Median : 0.00   Median : 0.000   Mode  :character   Median : 0.00
## Mean   : 58.01   Mean   : 7.449           Mean   : 65.46
## 3rd Qu.: 80.55   3rd Qu.: 0.000           3rd Qu.: 80.55
## Max.   :1001.18   Max.   :345.234           Max.   :1311.89
##
```

From 1975 to 2015 how many of storms have occurred each year and is the rate increasing?

```
ydf <- as.data.frame(table(df$year))
colnames(ydf) <- c('Year', 'Frequency')
ggplot(ydf, aes(x = Year, y = Frequency)) + geom_col(aes(fill = Frequency)) + geom_text(aes(label = Fre
```



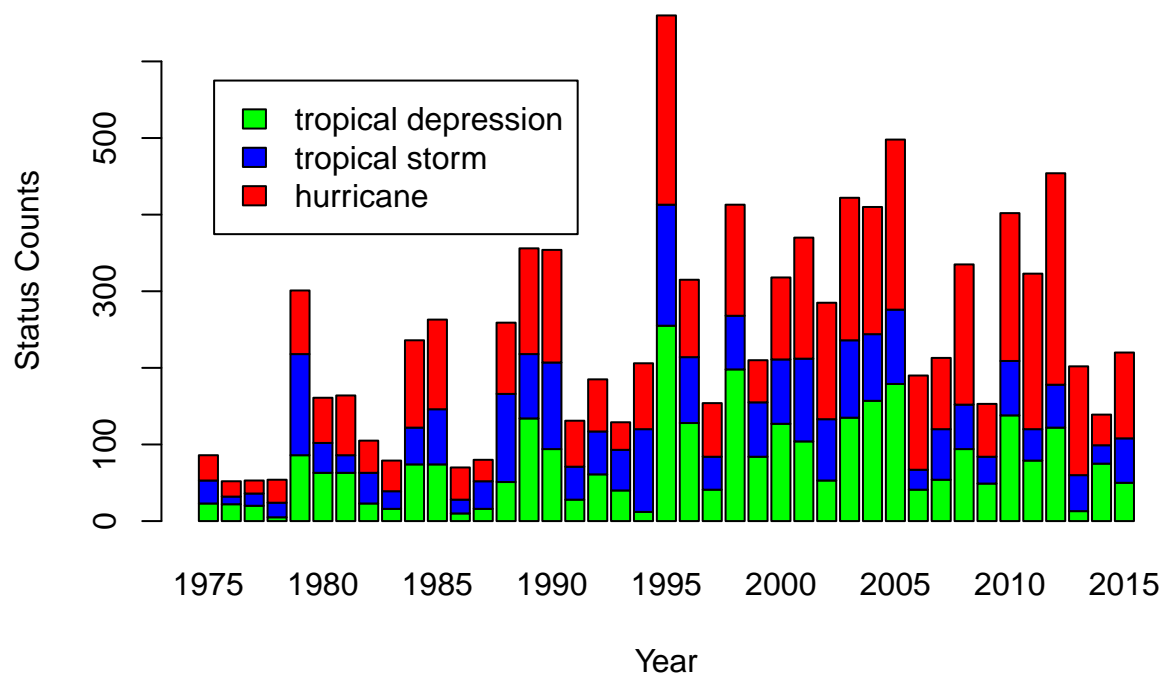
```
ydf$Year <- as.integer(ydf$Year)
ggplot(ydf, aes(x=Year, y=Frequency)) + geom_point(shape=18, color="blue") + geom_smooth(method=lm, linetype="dashed")
```



Both the bar graph and regression line show an increase in the number of storms for the last 40 years

How do the storms vary by category for each year?

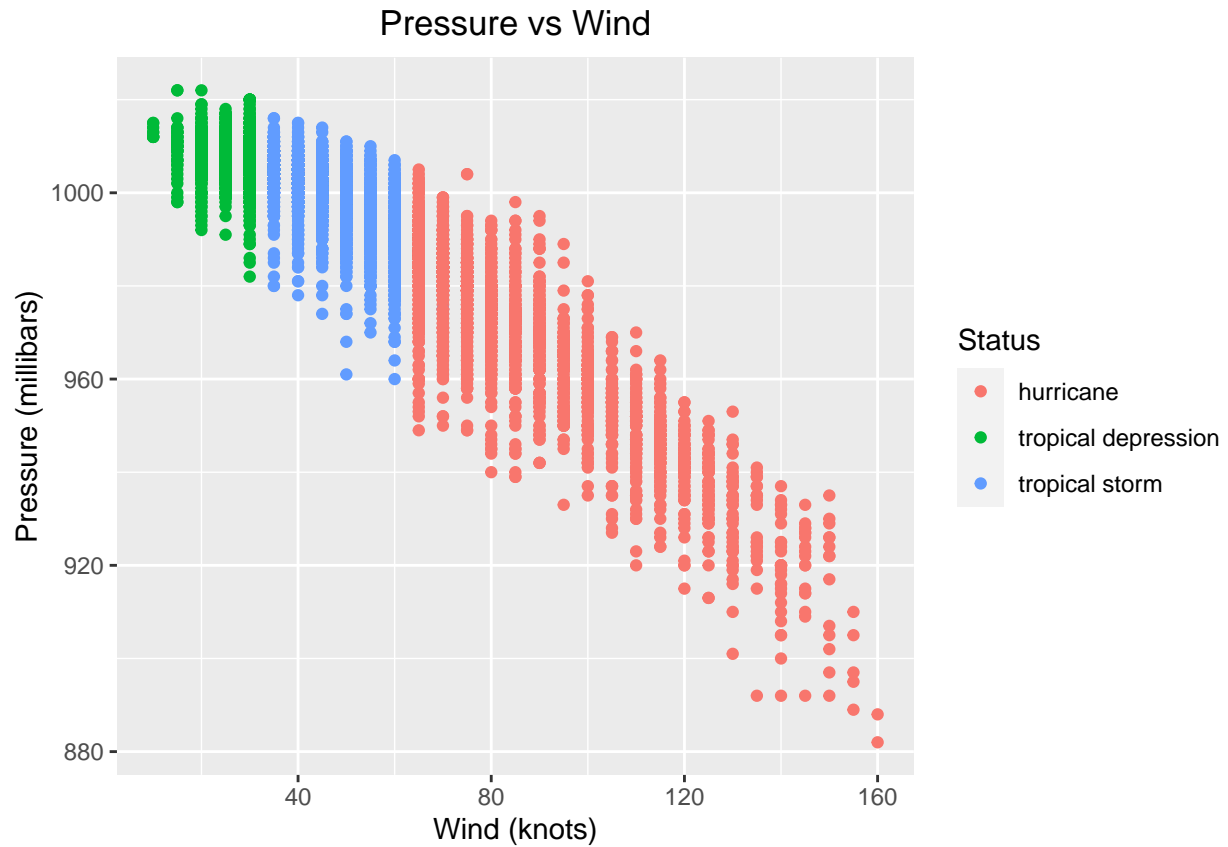
```
tb1 <- table(df$status, df$year)
barplot(tb1, xlab='Year', ylab='Status Counts', col=c('green','blue','red'))
legend(x=1, y=575, legend=unique(df$status), fill=c('green','blue','red'))
```



The rise of more storms has also lead to more hurricanes

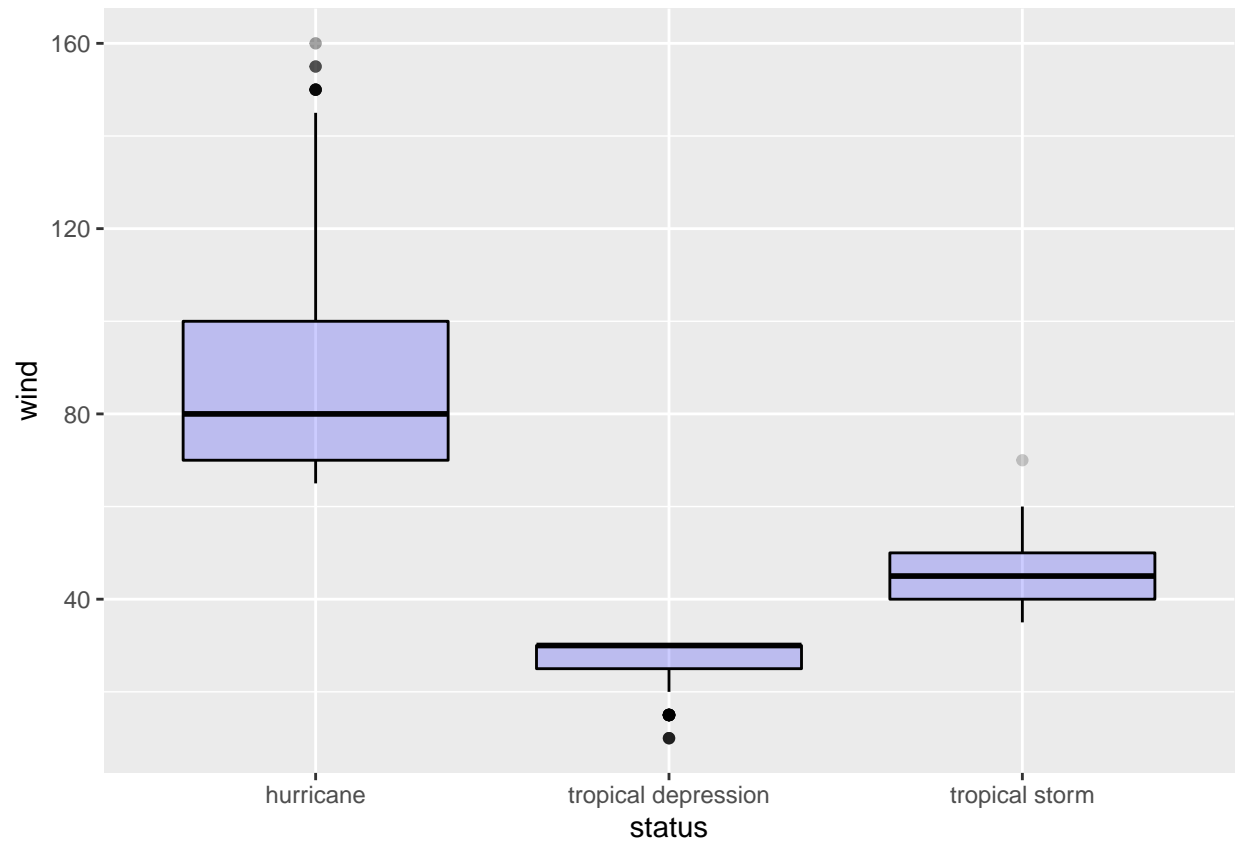
What makes a storm change status?

```
qplot(df$wind, df$pressure, main='Pressure vs Wind', xlab='Wind (knots)', ylab='Pressure (millibars)',
```



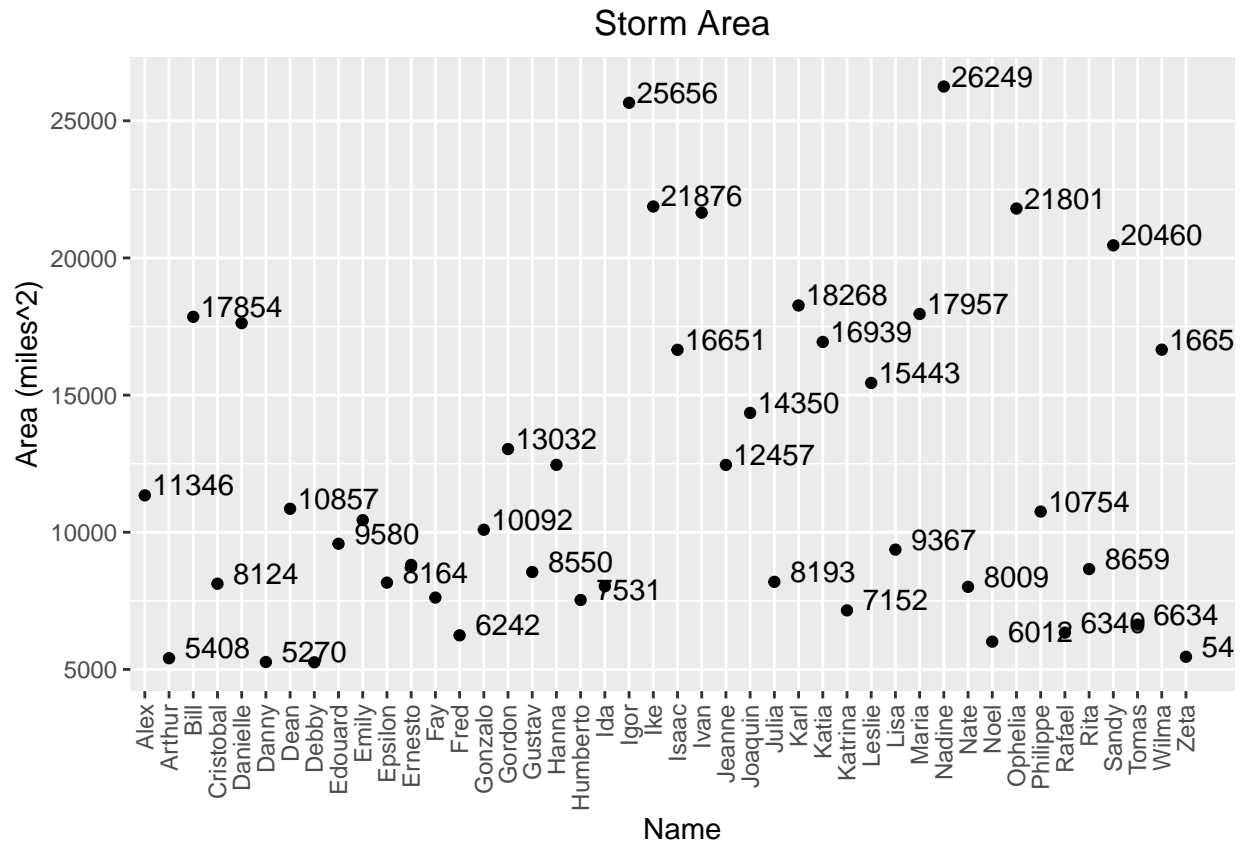
From the plot above it looks like there is a wide range in wind speed for a storm to be classified as a hurricane, that can be verified with a box plot

```
ggplot(df, aes(x=status, y=wind)) + geom_boxplot(color="black", fill="blue", alpha=0.2)
```



As expected hurricanes have the most values outside the box

Which storm has covered more land area ?



It looks like Nadine has covered more land than any other storm

Which storm name was most popular?

```
cdf <- count(df, name)
cdf <- filter(cdf, n > 70)
wordcloud(words = cdf$name, freq = cdf$n, color = 'blue', size = 1, shape = "rectangle", backgroundColo
```




The name Emily is used the most, let's take a look at her path and status change.

```
edf <- filter(df, name == 'Emily')
pal <- colorFactor(c('green', 'blue', 'red'), domain = c("tropical depression", "tropical storm", "hurricane"))

edf %>%
  leaflet(width = '100%') %>%
  addTiles() %>%
  setView(lng=-60, lat=32, zoom=3.3) %>%
  addCircleMarkers(lat = ~lat, lng = ~long, popup = edf$name, color=~pal(status), weight=2, stroke=FALSE)
```

PhantomJS not found. You can install it with `webshot::install_phantomjs()`. If it is installed, please

From the first plot we saw 1995 had the most storms, let's take a look at their pathing.

Conclusion

From 1975 to 2015 storms have affected many countries along the east cost including the US and Mexico. Since the current rate of storms are increasing we need to do more work in reducing human-made global warming.