

# Hurricanes EDA

## Team Members

Vipul Gharde - vig4

Animesh Sharma - as3592

Pavankumar Prakash Savanur - ps1149

Nikhil Panikulangara - npp88

---

## Data Sources

**amo\_by\_month.csv:**

[https://climatedataguide.ucar.edu/sites/default/files/amo\\_monthly.10yrLP.txt](https://climatedataguide.ucar.edu/sites/default/files/amo_monthly.10yrLP.txt)

(Manually generated CSV file from the textual data)

**category4\_atlantic\_by\_decade.csv:**

[https://en.wikipedia.org/wiki/List\\_of\\_Category\\_4\\_Atlantic\\_hurricanes](https://en.wikipedia.org/wiki/List_of_Category_4_Atlantic_hurricanes)

(Manually generated CSV file from the tables)

**category4\_atlantic\_by\_year.csv:**

[https://en.wikipedia.org/wiki/List\\_of\\_Category\\_4\\_Atlantic\\_hurricanes](https://en.wikipedia.org/wiki/List_of_Category_4_Atlantic_hurricanes)

(Manually generated CSV file from the tables)

**category5\_atlantic\_by\_decade.csv:**

[https://en.wikipedia.org/wiki/List\\_of\\_Category\\_5\\_Atlantic\\_hurricanes](https://en.wikipedia.org/wiki/List_of_Category_5_Atlantic_hurricanes)

(Manually generated CSV file from the tables)

**category5\_atlantic\_by\_year.csv:**

[https://en.wikipedia.org/wiki/List\\_of\\_Category\\_5\\_Atlantic\\_hurricanes](https://en.wikipedia.org/wiki/List_of_Category_5_Atlantic_hurricanes)

(Manually generated CSV file from the tables)

**tropical\_storm\_data.csv:**

[https://www.nhc.noaa.gov/TCR\\_StormReportsIndex.xml](https://www.nhc.noaa.gov/TCR_StormReportsIndex.xml)

(Converted XML to CSV File)

---

# Plot Bar Plots of Hurricanes by Decade

## Category 4 Hurricanes

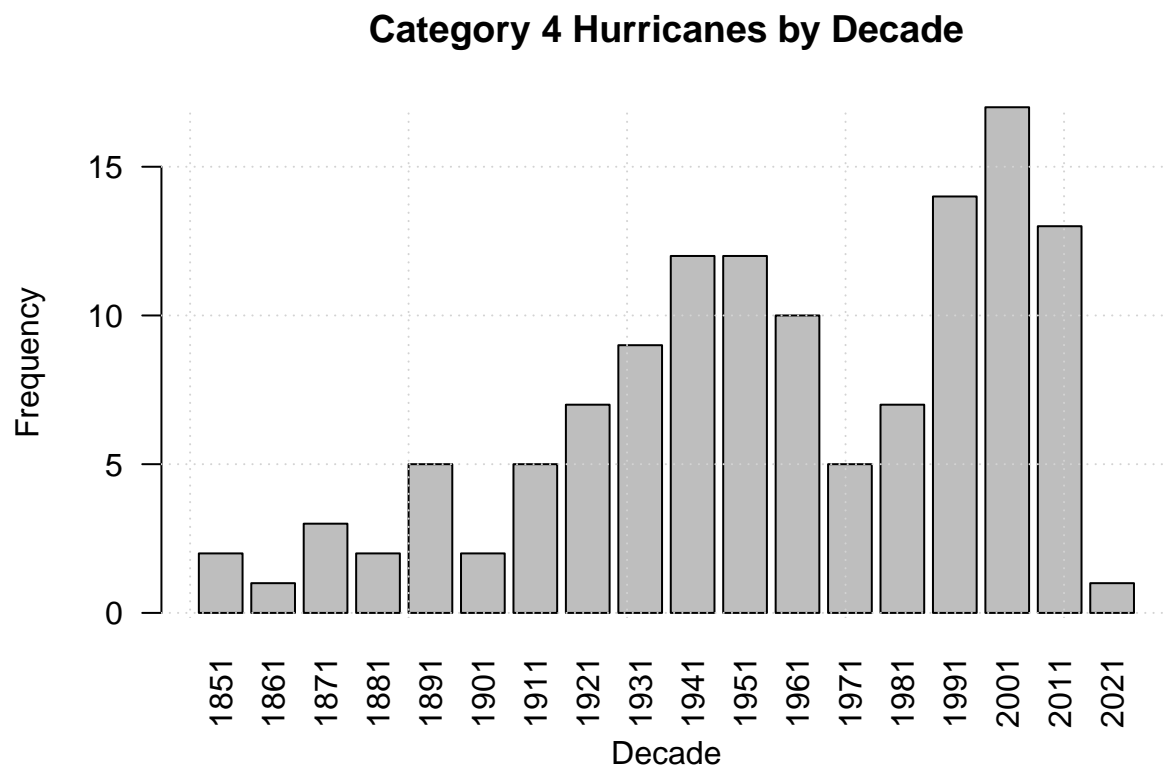
```
Cat4.Atlantic.decade <-  
  read.csv('datasets/category4_atlantic_by_decade.csv')  
Cat4.Atlantic.decade
```

##	decade	frequency
## 1	1851	2
## 2	1861	1
## 3	1871	3
## 4	1881	2
## 5	1891	5
## 6	1901	2
## 7	1911	5
## 8	1921	7
## 9	1931	9
## 10	1941	12
## 11	1951	12
## 12	1961	10
## 13	1971	5
## 14	1981	7
## 15	1991	14
## 16	2001	17
## 17	2011	13
## 18	2021	1

```

barplot(
  Cat4.Atlantic.decade$frequency,
  names.arg = Cat4.Atlantic.decade$decade,
  main = 'Category 4 Hurricanes by Decade',
  xlab = 'Decade',
  ylab = 'Frequency',
  las = 2
)
grid()

```



## Category 5 Hurricanes

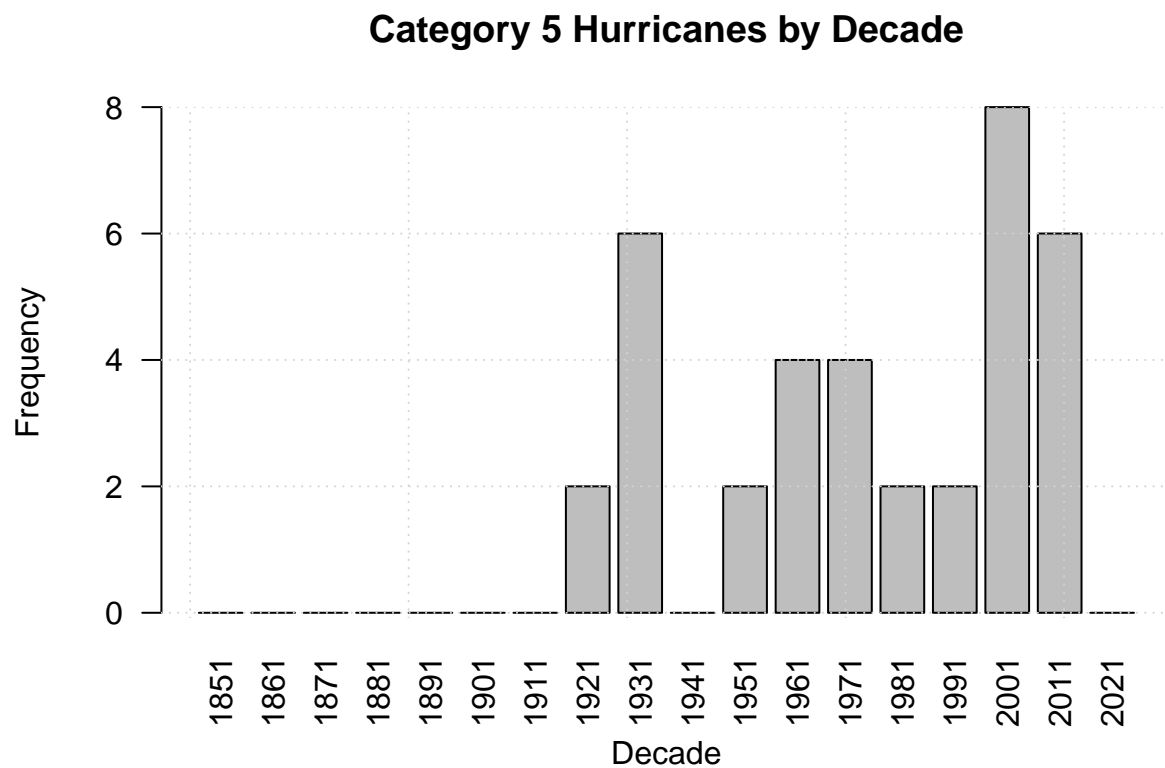
```
Cat5.Atlantic.decade <-  
  read.csv('datasets/category5_atlantic_by_decade.csv')  
Cat5.Atlantic.decade
```

##	decade	frequency
## 1	1851	0
## 2	1861	0
## 3	1871	0
## 4	1881	0
## 5	1891	0
## 6	1901	0
## 7	1911	0
## 8	1921	2
## 9	1931	6
## 10	1941	0
## 11	1951	2
## 12	1961	4
## 13	1971	4
## 14	1981	2
## 15	1991	2
## 16	2001	8
## 17	2011	6
## 18	2021	0

```

barplot(
  Cat5.Atlantic.decade$frequency,
  names.arg = Cat5.Atlantic.decade$decade,
  main = 'Category 5 Hurricanes by Decade',
  xlab = 'Decade',
  ylab = 'Frequency',
  las = 2
)
grid()

```



**Calculate the Highest Frequency of Hurricanes Over the Decades**

```

Cat4.max_freq = max(Cat4.Atlantic.decade$frequency)
Cat4.max_freq

```

```
## [1] 17
```

```

Cat5.max_freq = max(Cat5.Atlantic.decade$frequency)
Cat5.max_freq

```

```
## [1] 8
```

## Calculate Number of Decades

```
num_decades = length(Cat4.Atlantic.decade$decade)
num_decades
```

```
## [1] 18
```

## Function to Count the Number of Decades Having the Same Hurricane Frequencies

Example: Number of Decades where there was 1 occurrence of Category 4 Hurricane is 2,  
Number of Decades where there were 2 occurrences of Category 4 Hurricanes is 3, etc.

```
get_freqs <- function(Cat) {
  freqs = vector(mode = 'integer', length = max(Cat$frequency) + 1)

  for (i in Cat$frequency) {
    freqs[i + 1] = freqs[i + 1] + 1
  }
  return(freqs)
}
```

```
Cat4.freqs = get_freqs(Cat4.Atlantic.decade)
Cat4.freqs
```

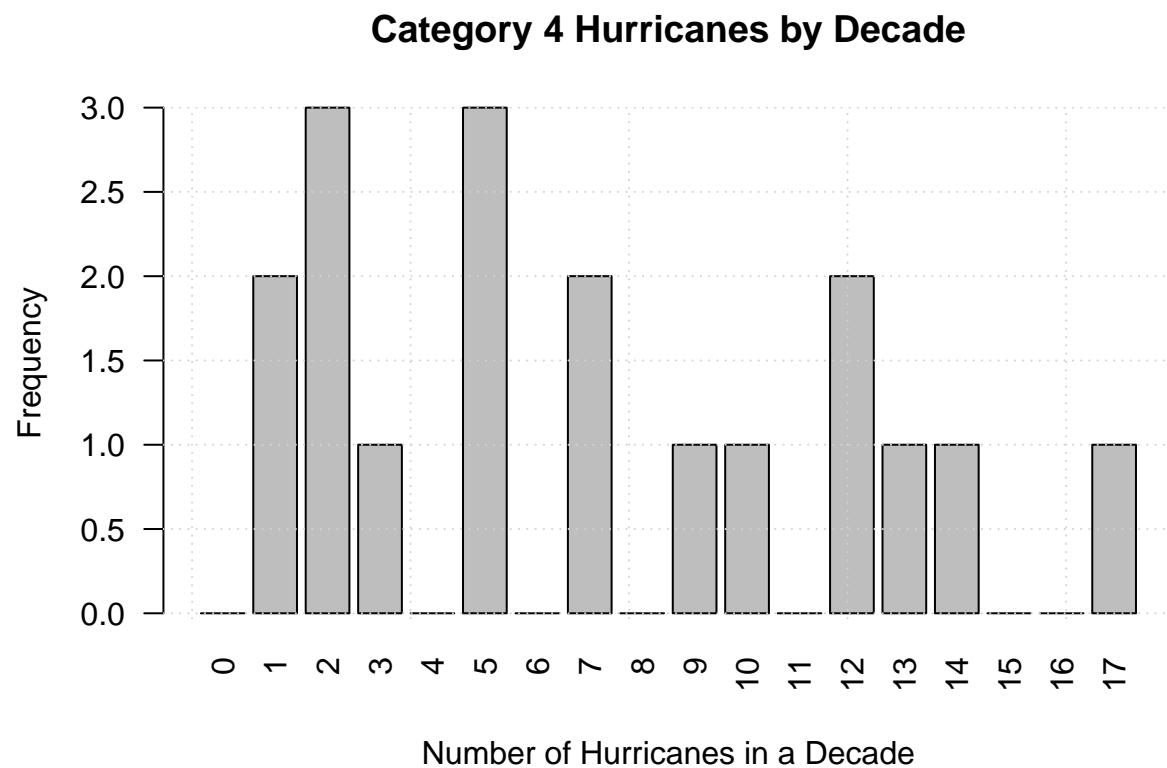
```
## [1] 0 2 3 1 0 3 0 2 0 1 1 0 2 1 1 0 0 1
```

```
Cat5.freqs = get_freqs(Cat5.Atlantic.decade)
Cat5.freqs
```

```
## [1] 9 0 4 0 2 0 2 0 1
```

## Plot Number of Hurricanes in a Decade

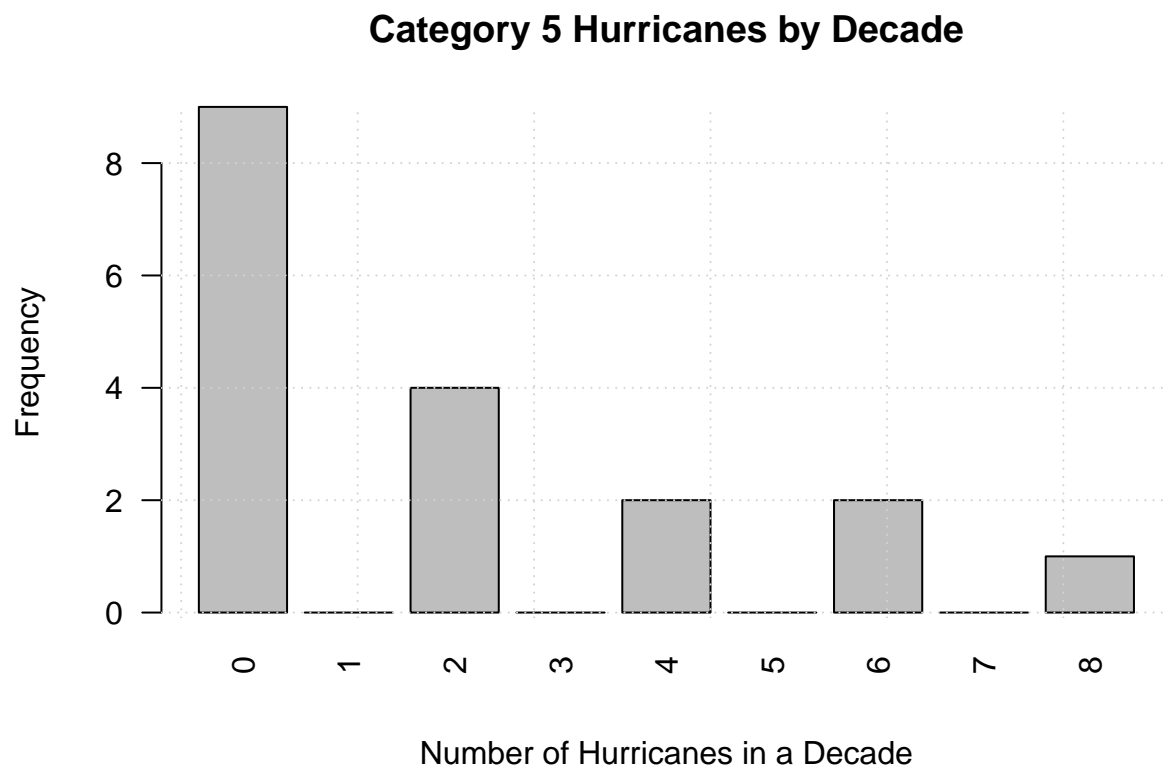
```
barplot(  
  Cat4.freqs,  
  names.arg = 0:Cat4.max_freq,  
  main = 'Category 4 Hurricanes by Decade',  
  xlab = 'Number of Hurricanes in a Decade',  
  ylab = 'Frequency',  
  las = 2  
)  
grid()
```



```

barplot(
  Cat5.freqs,
  names.arg = 0:Cat5.max_freq,
  main = 'Category 5 Hurricanes by Decade',
  xlab = 'Number of Hurricanes in a Decade',
  ylab = 'Frequency',
  las = 2
)
grid()

```



## Calculate lambda Value for Poisson Distribution

```

Cat4.lambda = mean(Cat4.Atlantic.decade$frequency)
Cat4.lambda

```

```
## [1] 7.055556
```

```

Cat5.lambda = mean(Cat5.Atlantic.decade$frequency)
Cat5.lambda

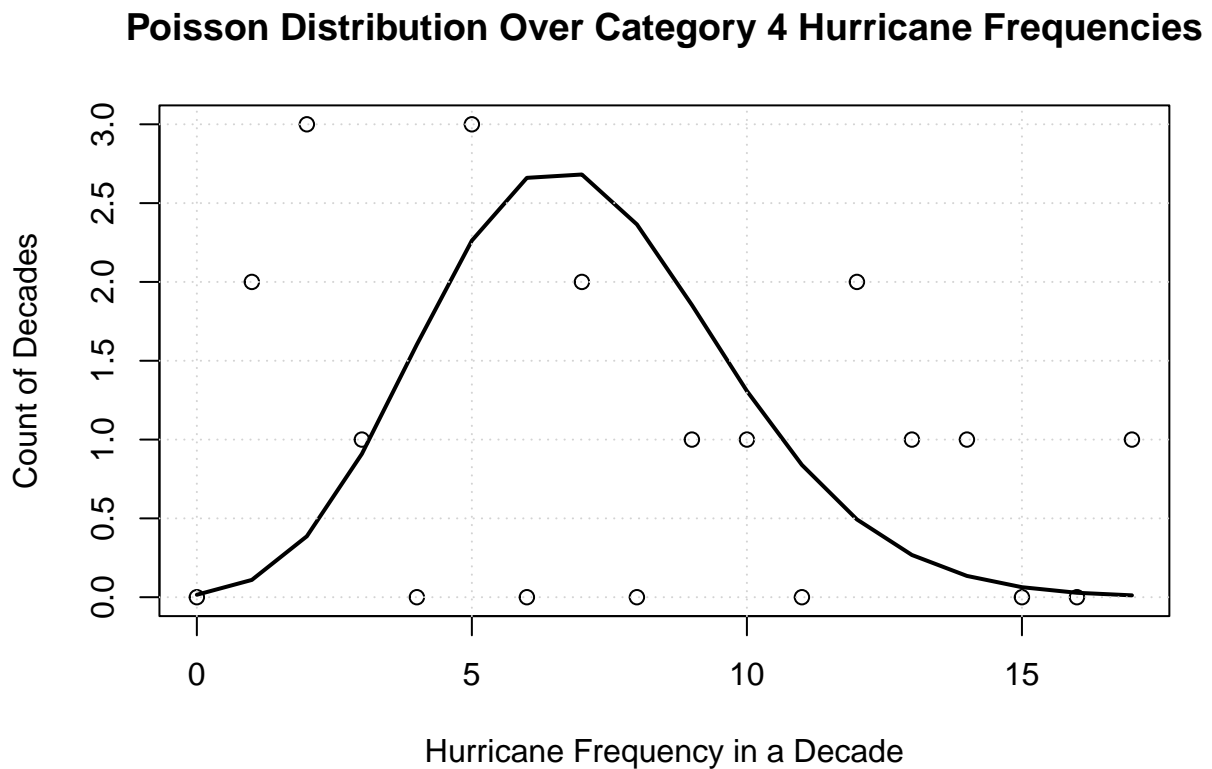
```

```
## [1] 2
```



## Plot Poisson Distribution Over Hurricane Frequencies

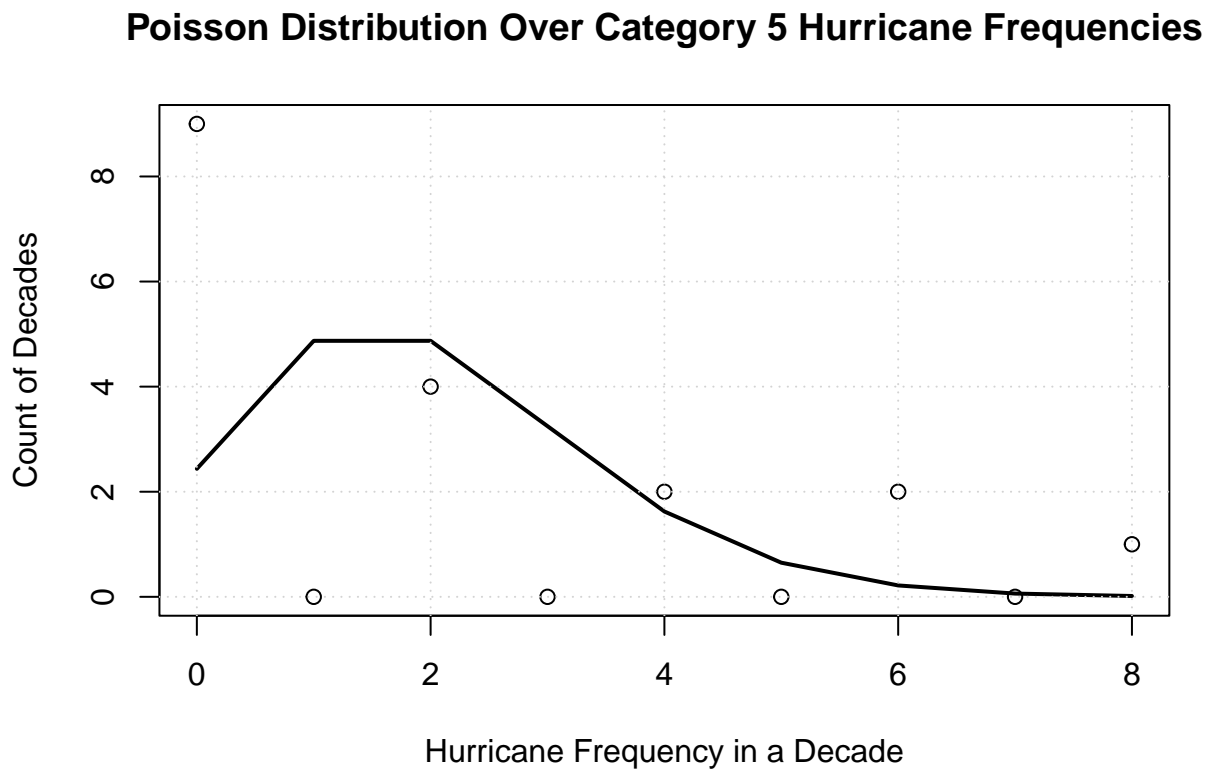
```
plot(
  0:Cat4.max_freq,
  Cat4.freqs,
  main = 'Poisson Distribution Over Category 4 Hurricane Frequencies',
  xlab = 'Hurricane Frequency in a Decade',
  ylab = 'Count of Decades'
)
lines(0:Cat4.max_freq,
      dpois(0:Cat4.max_freq, Cat4.lambda) * sum(Cat4.freqs),
      lwd = 2)
grid()
```



```

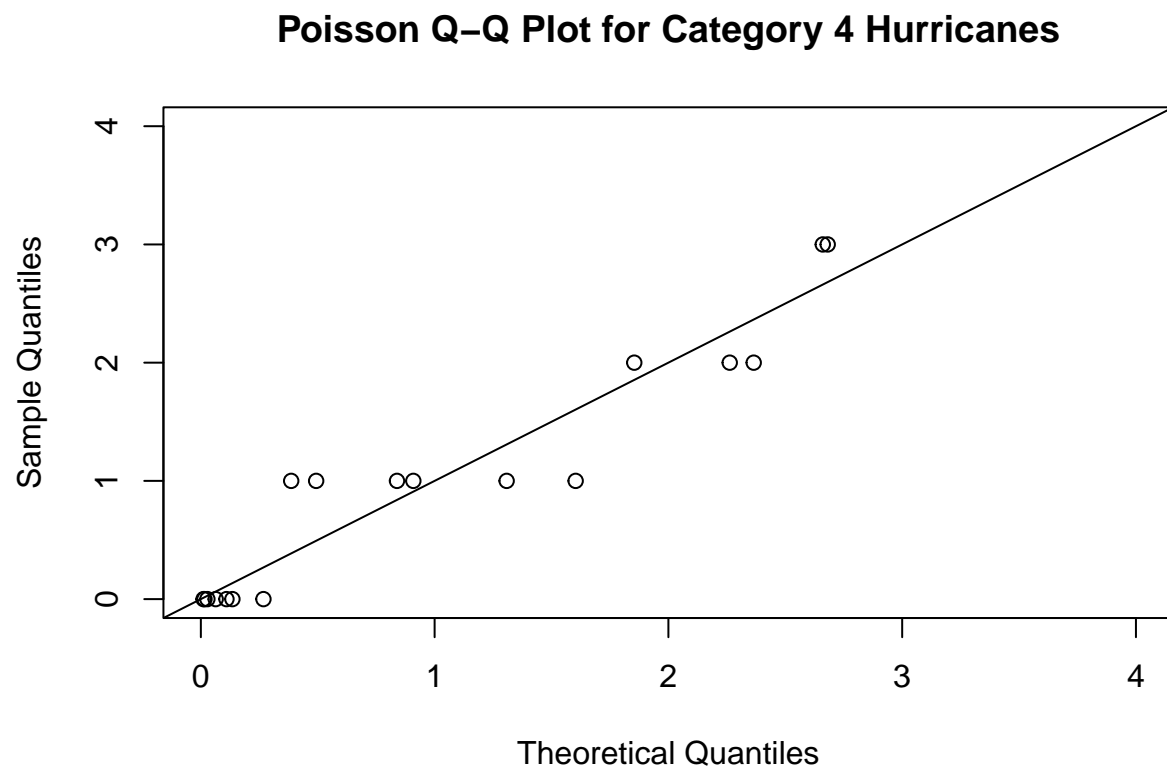
plot(
  0:Cat5.max_freq,
  Cat5.freqs,
  main = 'Poisson Distribution Over Category 5 Hurricane Frequencies',
  xlab = 'Hurricane Frequency in a Decade',
  ylab = 'Count of Decades'
)
lines(0:Cat5.max_freq,
      dpois(0:Cat5.max_freq, Cat5.lambda) * sum(Cat5.freqs),
      lwd = 2)
grid()

```

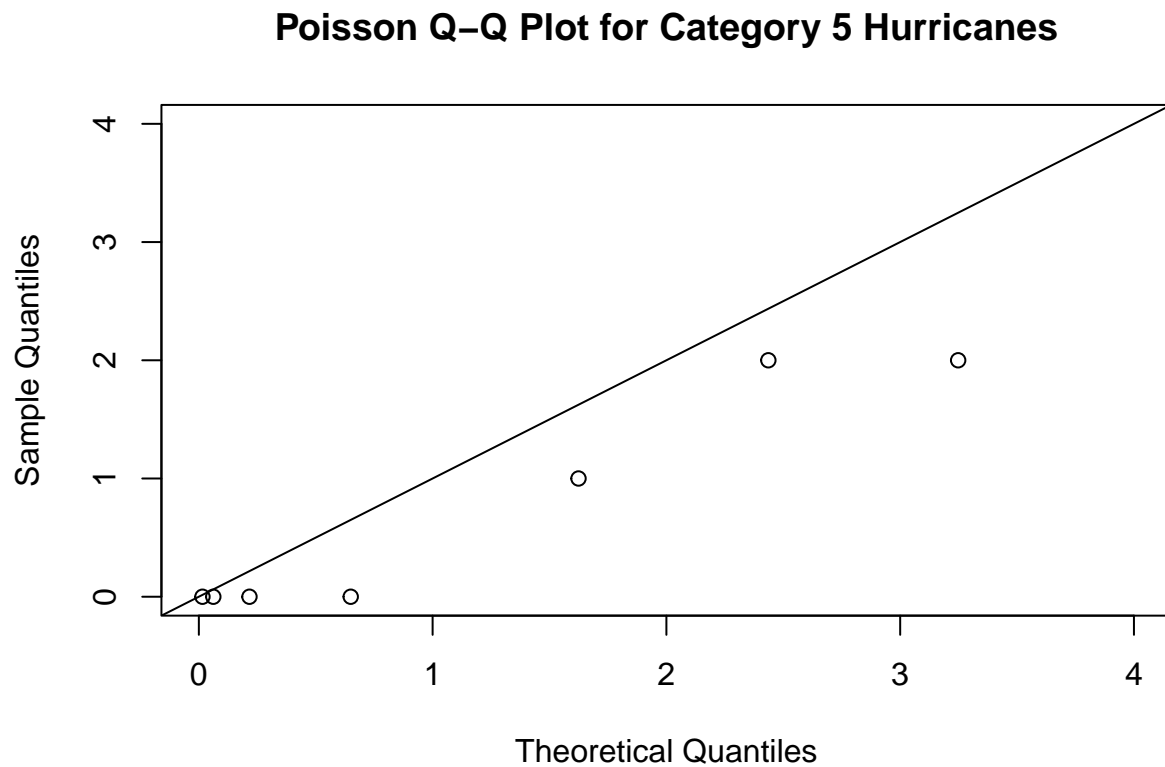


## Plot Q-Q Plots

```
qqplot(  
  dpois(0:Cat4.max_freq, Cat4.lambda) * sum(Cat4.freqs),  
  Cat4.freqs,  
  main = 'Poisson Q-Q Plot for Category 4 Hurricanes',  
  xlab = 'Theoretical Quantiles',  
  ylab = 'Sample Quantiles',  
  xlim = c(0, 4),  
  ylim = c(0, 4)  
)  
abline(0, 1)
```



```
qqplot(
  dpois(0:Cat5.max_freq, Cat5.lambda) * sum(Cat5.freqs),
  Cat5.freqs,
  main = 'Poisson Q-Q Plot for Category 5 Hurricanes',
  xlab = 'Theoretical Quantiles',
  ylab = 'Sample Quantiles',
  xlim = c(0, 4),
  ylim = c(0, 4)
)
abline(0, 1)
```



## Read Monthly Atlantic Multidecadal Oscillation Index CSV File

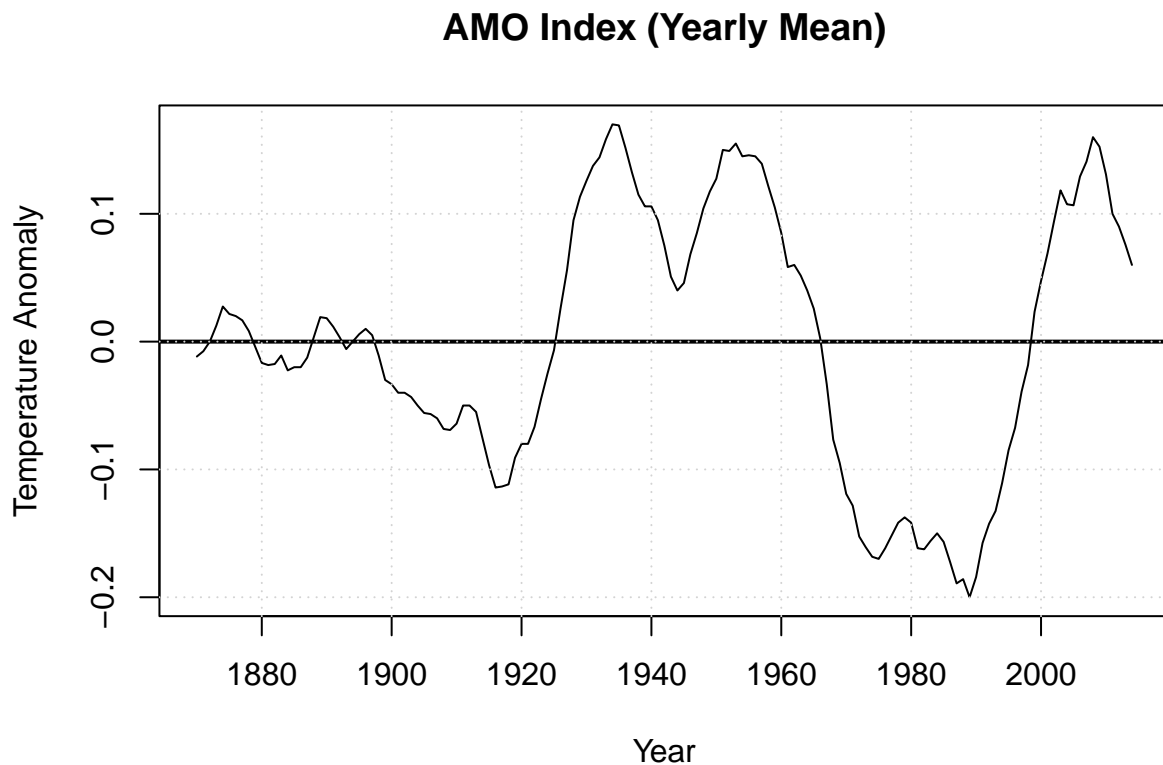
```
AMO.month <- read.csv('datasets/amo_by_month.csv')
AMO.month
```

## Remove Outlier Values

```
AMO.month = AMO.month[c(-146:-152), ]
AMO.month
```

## Plot AMO Index (Yearly Mean)

```
AMO.month.means = rowMeans(AMO.month[, -1])
AMO.month.means.ts = ts(AMO.month.means, frequency = 1, start = 1870)
plot(AMO.month.means.ts,
     main = 'AMO Index (Yearly Mean)',
     xlab = 'Year',
     ylab = 'Temperature Anomaly')
abline(h = 0, lwd = 2)
grid()
```

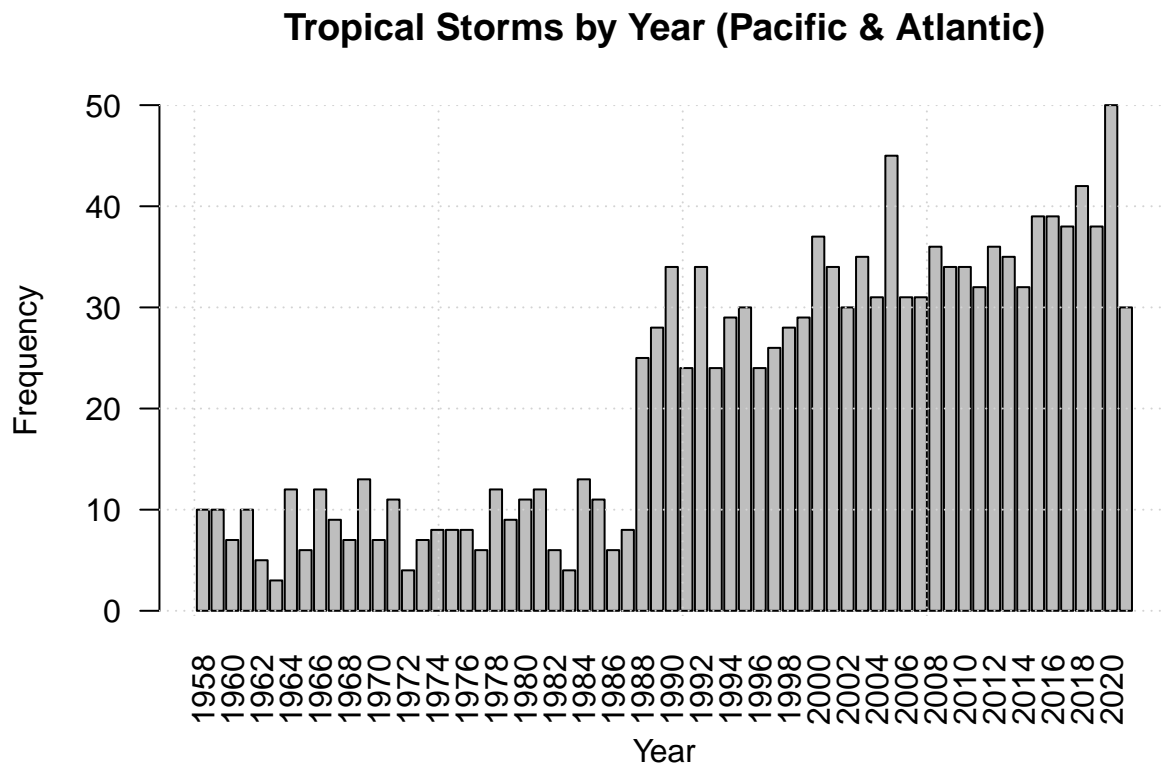


## Read Annual Tropical Storm Data from 1958 - 2021

```
TS <- read.csv('datasets/tropical_storm_data.csv')
TS
```

## Plot a Bar Plot with Year on x-axis and Storm Frequency on y-axis

```
barplot(
  as.data.frame(table(TS$Year))$Freq,
  names.arg = unique(TS$Year),
  main = 'Tropical Storms by Year (Pacific & Atlantic)',
  xlab = 'Year',
  ylab = 'Frequency',
  las = 2
)
grid()
```



## Calculate the Highest Frequency of Tropical Storms Over the Years

```
TS.max_freq = max(as.data.frame(table(TS$Year))$Freq)
TS.max_freq
```

```
## [1] 50
```

## Count the Number of Years Having the Same Tropical Storm Frequencies

Example: Number of Years where there were 3 occurrences of Tropical Storm is 1,  
Number of Years where there were 4 occurrences of Tropical Storm is 2, etc.

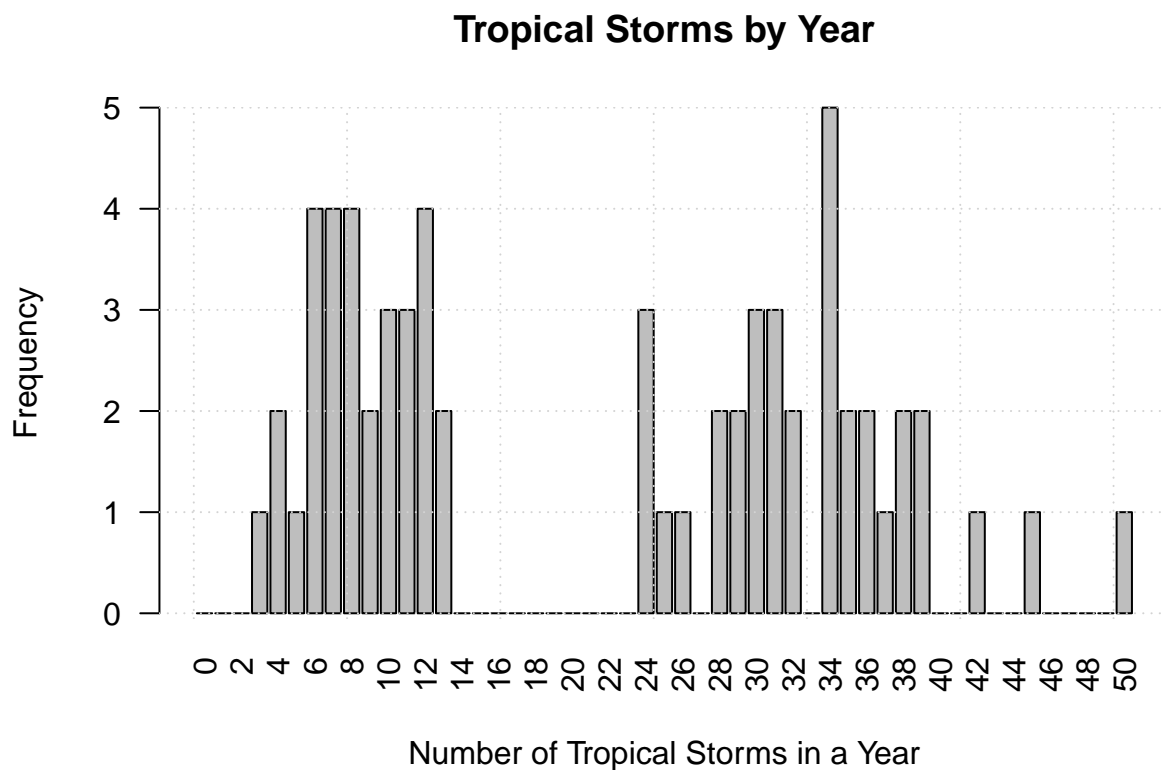
```
TS.freqs = vector(mode = 'integer', length = max(as.data.frame(table(TS$Year))$Freq) + 1)
for (i in as.data.frame(table(TS$Year))$Freq) {
  TS.freqs[i + 1] = TS.freqs[i + 1] + 1
}

TS.freqs
```

```
## [1] 0 0 0 1 2 1 4 4 4 2 3 3 4 2 0 0 0 0 0 0 0 0 3 1 1 0 2 2 3 3 2 0 5 2 2 1
## [39] 2 2 0 0 1 0 0 1 0 0 0 0 1
```

## Plot Number of Tropical Storms in a Year

```
barplot(  
  TS.freqs,  
  names.arg = 0:TS.max_freq,  
  main = 'Tropical Storms by Year',  
  xlab = 'Number of Tropical Storms in a Year',  
  ylab = 'Frequency',  
  las = 2  
)  
grid()
```



## Calculate lambda Value for Poisson Distribution

```
TS.lambda = mean(as.data.frame(table(TS$Year))$Freq)  
TS.lambda
```

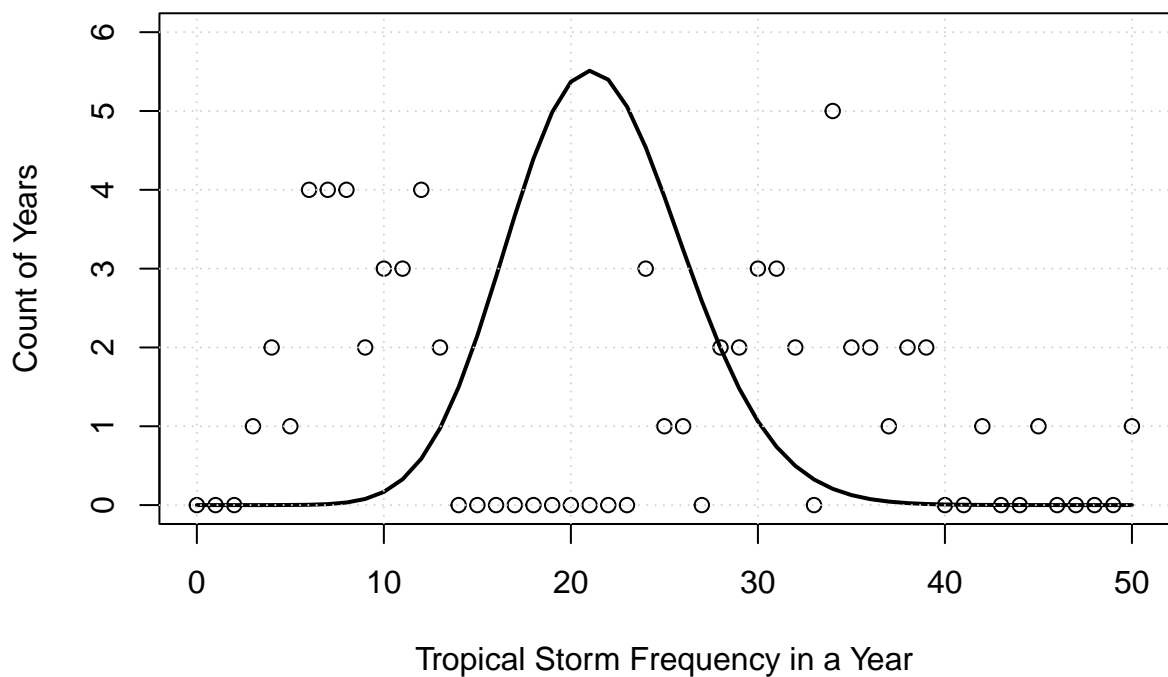
```
## [1] 21.54688
```



## Plot Poisson Distribution Over Tropical Storm Frequencies

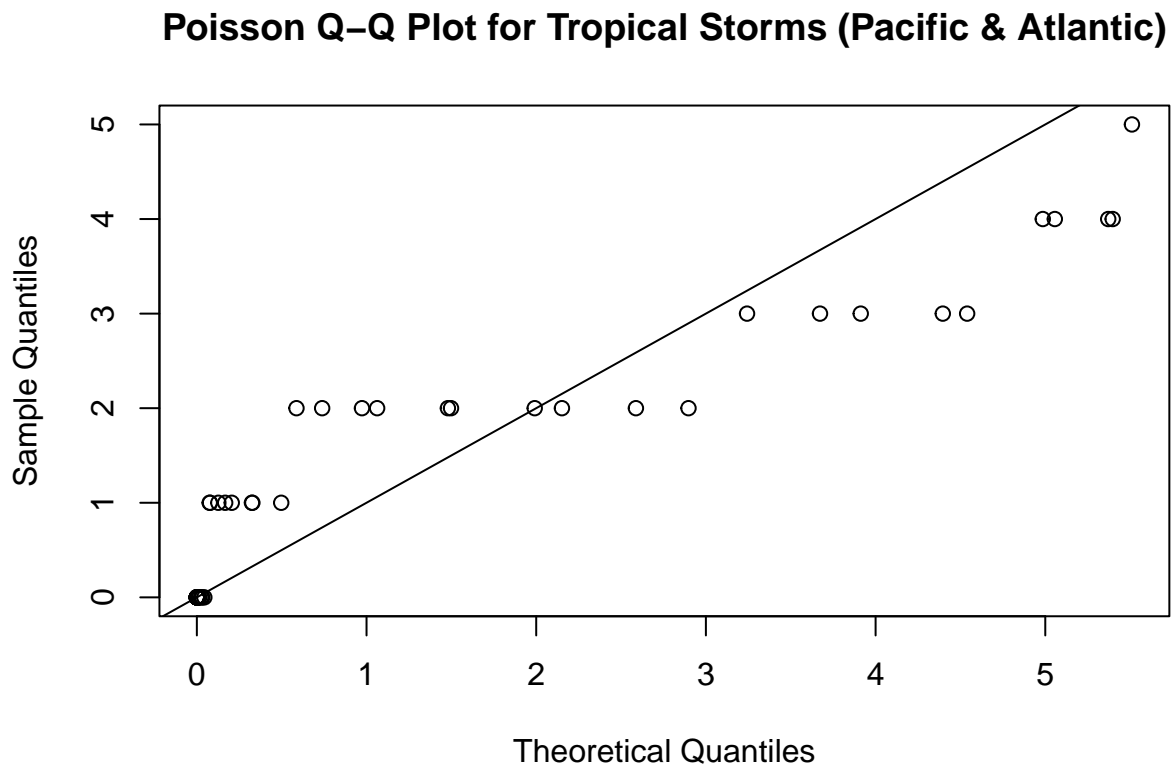
```
plot(
  0:TS.max_freq,
  TS.freqs,
  main = 'Poisson Distribution over Tropical Storm (Pacific & Atlantic) Frequencies',
  xlab = 'Tropical Storm Frequency in a Year',
  ylab = 'Count of Years',
  ylim = c(0, 6)
)
lines(0:TS.max_freq,
      dpois(0:TS.max_freq, TS.lambda) * sum(TS.freqs),
      lwd = 2)
grid()
```

### Poisson Distribution over Tropical Storm (Pacific & Atlantic) Frequencies



## Plot the Q-Q Plot

```
qqplot(
  dpois(0:TS.max_freq, TS.lambda) * sum(TS.freqs),
  TS.freqs,
  main = 'Poisson Q-Q Plot for Tropical Storms (Pacific & Atlantic)',
  xlab = 'Theoretical Quantiles',
  ylab = 'Sample Quantiles'
)
abline(0, 1)
```



## Conclusion

The Tropical Storm dataset contains data for Tropical Depressions, Tropical Storms and Hurricanes for both East Pacific and Atlantic Oceans. As such, Tropical Storms as a whole are not as rare as Hurricanes. Therefore, the Q-Q Plot for the Poisson Distribution of the data doesn't properly fit, unlike the isolated Hurricane datasets.