Analysis of Atlantic Hurricane Seasons, 1851-2020

Christopher Rutherford

Fair warning: I'm not a meteorologist or atmospheric scientist. This is for conducting some basic EDA for fun and practicing some R, not proving or predicting anything regarding the climate.

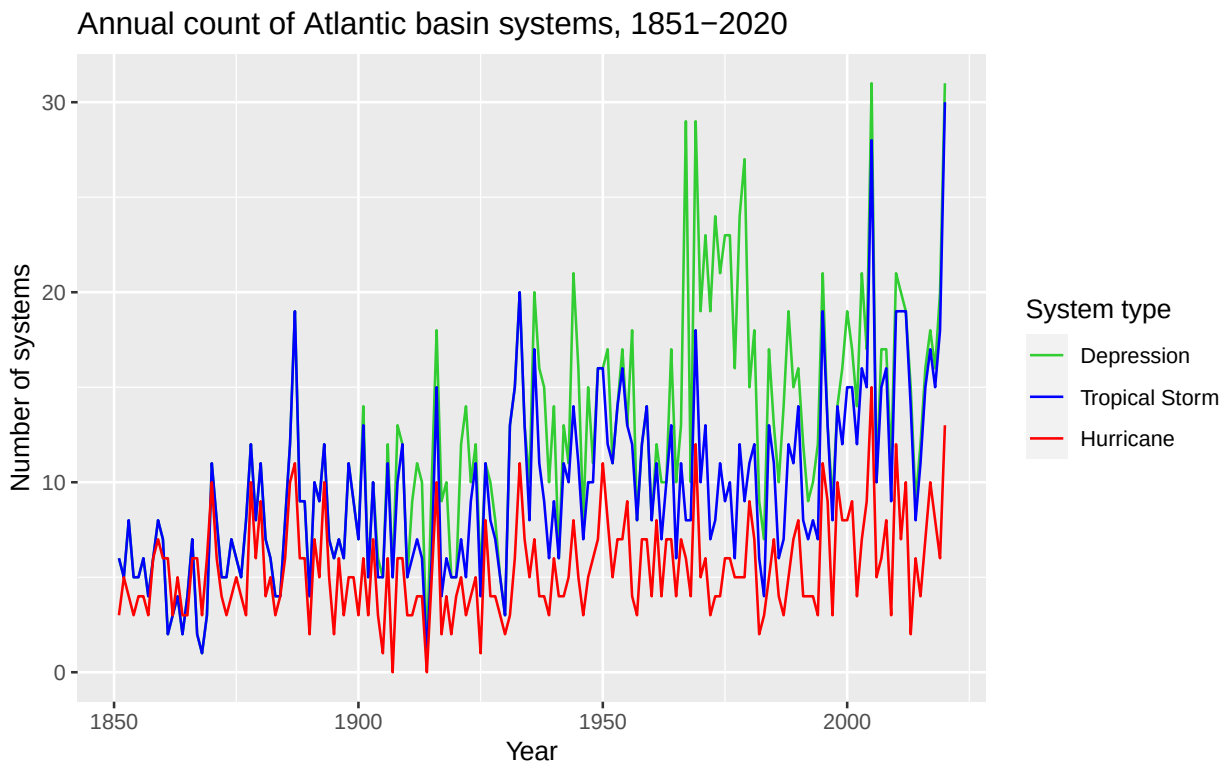
In this project, we'll be performing some exploratory analysis on Atlantic hurricane seasons since 1851. Since reliable record keeping started this year, data for previous years is very sparse and typically includes only those that made landfall.

```
library(readr)
hurricanes <- read_csv("hurricanes.csv",
                       col_types = cols(Year = col_date(format = "%Y")))
colnames(hurricanes) <- c("Year", "TotalSystems", "TropicalStorms",
                          "Hurricanes", "MajorHurricanes", "ACE")
hurricanes
```

```
## # A tibble: 170 x 6
##    Year       TotalSystems TropicalStorms Hurricanes MajorHurricanes   ACE
##    <date>            <dbl>          <dbl>      <dbl>           <dbl> <dbl>
##  1 1851-01-01            6              6          3               1    36
##  2 1852-01-01            5              5          5               1    73
##  3 1853-01-01            8              8          4               2    76
##  4 1854-01-01            5              5          3               1    31
##  5 1855-01-01            5              5          4               1    18
##  6 1856-01-01            6              6          4               2    49
##  7 1857-01-01            4              4          3               0    40
##  8 1858-01-01            6              6          6               0    45
##  9 1859-01-01            8              8          7               1    56
## 10 1860-01-01            7              7          6               1    62
## # ... with 160 more rows
```

We can take a look at how the amount of depressions, storms, and hurricanes has changed over the years.

```
system_counts <- hurricanes[, 1:4]
count_melt <- melt(system_counts, id = "Year")
ggplot(count_melt, aes(x = Year, y = value, color = variable))+
  geom_line()+
  labs(title = "Annual count of Atlantic basin systems, 1851-2020",
       x = "Year", y = "Number of systems")+
  scale_color_manual(values = c("limegreen", "blue", "red"),
                     name = "System type",
                     labels = c("Depression", "Tropical Storm", "Hurricane"))
```

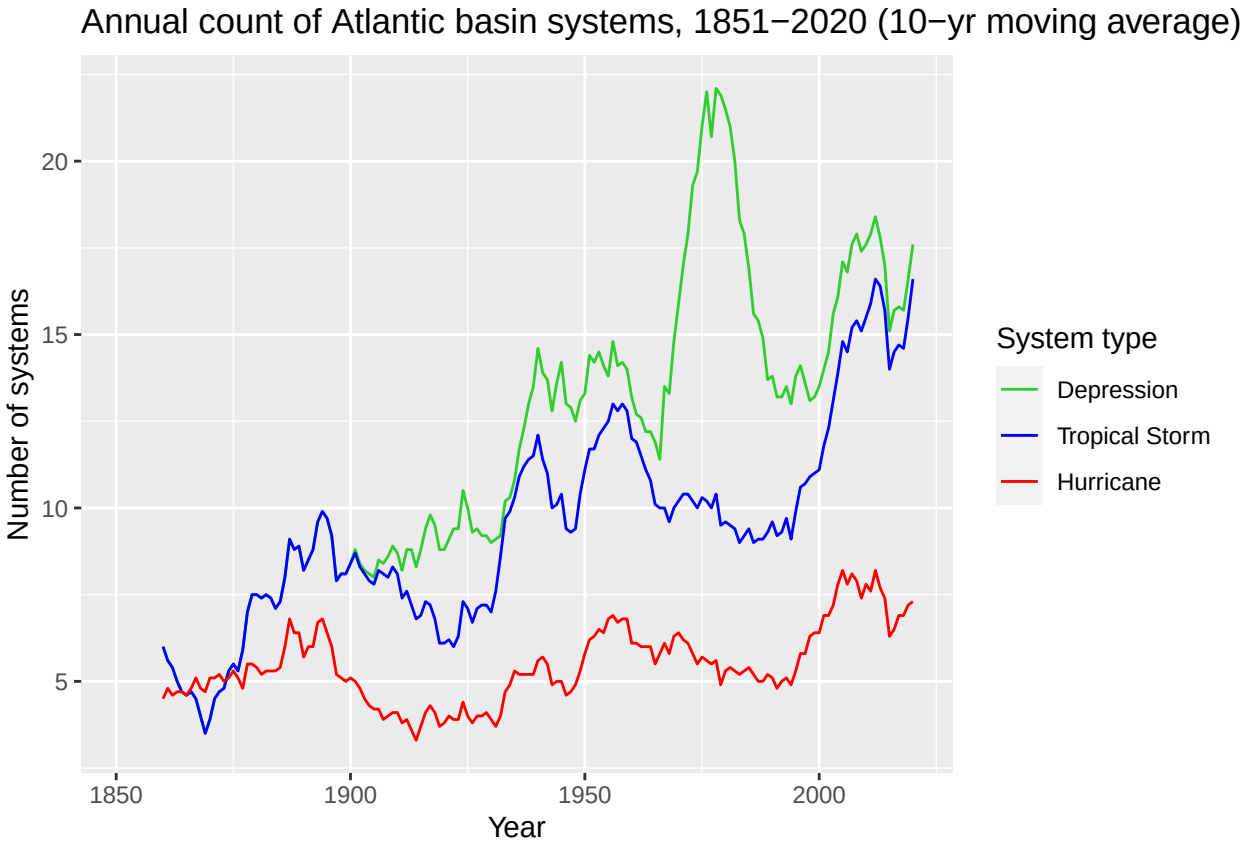


Due to the lack of technology capable of detecting hurricanes far from land, the amount of total systems is equal to the amount of tropical storms for the first 50 or so years. We may be interested in using moving averages for this data due to the year-to-year noise observed. We'll use simple moving averages to get an idea of the overall trend over the years.

```
hurricanes = hurricanes %>%
  mutate(sma_td = SMA(TotalSystems),
         sma_ts = SMA(TropicalStorms),
         sma_hur = SMA(Hurricanes))
sma_all = hurricanes[, c(1, 7:9)]
sma_melt = melt(sma_all, id = "Year")
```

Now that we've created and melted our moving averages, we can plot them.

```
ggplot(sma_melt, aes(x = Year, y = value, color = variable))+
  geom_line(na.rm = T)+
  labs(title = "Annual count of Atlantic basin systems, 1851-2020 (10-yr moving average)",
       x = "Year", y = "Number of systems")+
  scale_color_manual(values = c("limegreen", "blue", "red"),
                     name = "System type",
                     labels = c("Depression", "Tropical Storm", "Hurricane"))
```



While this plot does remove most of the spikes of the previous one, helps paint a much clearer picture of how trends have changed overtime. Most notably, there are periods of time where more hurricanes form throughout the season, and these periods are followed by periods of decreased activity. What really stands out, however, is the sudden spike in the number of tropical depressions in the 1970s. This is likely due to the satellite era commencing in 1967, making it much easier to detect storms that are either very weak or remain far from land.

Another important variable to consider beyond cyclone count is accumulated cyclone energy (ACE). With ACE, we can get a better idea of how powerful and active a season is. A higher ACE typically indicates more intense cyclones. As such, seasons with less storms can have a higher ACE than those with more storms.

```
count_ace = hurricanes[,c(1:4, 6)]
count_ace_melt = melt(count_ace, id = "Year")
sma_ace = hurricanes[,c(1, 6:9)]
sma_ace_melt = melt(sma_ace, id = "Year")
```

Since not using the moving averages makes the plot messy, we will focus on a set of particularly active years (2004, 2005) along with a quieter season with a single strong storm (1992, Andrew).

```
ggplot(hurricanes, aes(x = Year))+
  geom_line(aes(y = TotalSystems, color = "Depression"), na.rm = T)+
  geom_line(aes(y = TropicalStorms, color = "Tropical Storm"), na.rm = T)+
  geom_line(aes(y = Hurricanes, color = "Hurricane"),  na.rm = T)+
  labs(title = "Annual count and ACE of Atlantic hurricane seasons, 1851-2020",
       x = "Year", color = "Key")+
```

```
scale_y_continuous(name = "System Count",
                   sec.axis = sec_axis(trans = ~.*10,
                                       name = "ACE"))+
geom_line(aes(y = ACE/10, color = "ACE"), na.rm = T, lty = 5)+
scale_x_date(limits = as.Date(c("1992-01-01", "2006-01-01")))+
theme(legend.position = c(0.12, 0.81))
```