

# Climatology of Atlantic Hurricanes

*Dean D. Churchill, Ph.D.*

*August 13, 2014*

## Abstract

I describe the climatology of tropical cyclonic storms over the Atlantic Ocean basin. Key goals of the study are to understand what portions of the United States are most at risk from damaging hurricanes, and whether or not that risk is changing as the climate changes.

## Introduction

The National Hurricane Center maintains a comprehensive data base of tropical storms and hurricanes. The database is downloaded from this url: <http://www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2013-052714.txt> (<http://www.nhc.noaa.gov/data/hurdat/hurdat2-1851-2013-052714.txt>) This HURDAT2 file contains all records of storms in the Atlantic from 1851 through 2013. The format of this data set is documented at <http://www.nhc.noaa.gov/data/hurdat/hurdat2-format-nencpac.pdf> (<http://www.nhc.noaa.gov/data/hurdat/hurdat2-format-nencpac.pdf>)

I am exploring this data looking for trends and characteristics of where storms form, how the number and intensity change over time, and whether there is any climate change signal evident in the data. These are important questions to ask as the economic impact, not the mention the death and injuries that come from hurricanes, is a big issue for the United States and countries of Central America and the Caribbean.

The first step in analyzing this data is to convert the track data from text format to a R dataframe. The `setup()` function performs this process. Setup also loads the libraries that are needed for this project.

The parsing process assumes that the `hurdat2` file, referenced above, is stored in the directory `~/hurricanes`. The `setup` function will parse the hurricane data and store to the file `~/hurricanes/hurtrack.Rda`.

A record of data is recorded once every 6 hours for each storm. The record is comprised of the date and time, storm identifier, latitude and longitude of the center of the storm, the maximum wind speed (in knots) observed in the storm and the minimum pressure (in millibars). The National Hurricane Center defines the type of storm based on the maximum wind speed observed every 6 hours. Tropical depressions (TD), and tropical storms (TS) are the weak storms in terms of wind. Hurricanes have 5 categories of intensity, labeled below as 'Cat 1' (the weakest) through 'Cat 5' (the most intense). During the parsing I used the `cut` function to categorize the storms, and I will use the resulting 'Cat' field during analysis.

First step is to set up the libraries and parse the data if needed. The first time this process runs, the data will be parsed and stored as a dataframe. The run time is about 2 minutes for this process to complete. If all the data was read in, there will be 1771 storms in the data frame. The function `readstorm()` does the reading and parsing of data.

```

#
# readStorm - reads in all the storm track data from a file connection
# written 7.17.14 by Dean Churchill
# input: connection to an open file with hurricane track data in it
# output: a dataframe with the storm data in it
#
readStorm <- function(con) {

  header <- readLines(con, n=1) # read in the header record
  fields <- strsplit(header, ",") # parse on commas
  uid <- fields[[1]][1] # unique storm id.
  basin <- substr(uid, 1, 2) # AL always
  stormid <- substr(uid, 3, 4)
  storm_year <- as.numeric(substr(uid, 5, 8)) # year of the storm

  name <- gsub(" ", "", fields[[1]][2]) # name of the storms
  nrows <- as.integer(fields[[1]][3]) # number of rows of track data

  text <- readLines(con, n=nrows) # read all lines of text in report
  lrec <- 1:nrows # add vector of logical record number
  text <- gsub(" ", "", text) # remove whitespace before parsing
  t <- strsplit(text, ",") # parse the track data
  # converts to a matrix format for referencing as vector operations
  t <- do.call(rbind, t)
  # extract date and time of observation, GMT.

  year <- as.integer(substr(t[,1], 1, 4))
  month <- as.integer(substr(t[,1], 5, 6))
  day <- as.integer(substr(t[,1], 7, 8))
  hour <- as.integer(substr(t[,2], 1, 2))
  minute <- as.integer(substr(t[,2], 3, 4))
  # create a date object for use in analysis

  dtg <- ISOdatetime(year, month, day, hour, minute, 0, tz='GMT')

  rec_type <- t[,3] # identifies the type of record
  # rec_type
  #L - Landfall (center of system crossing a coastline)
  #W - Maximum sustained wind speed
  #P - Minimum in central pressure
  #I - An intensity peak in terms of both pressure and wind
  #C - Closest approach to a coast, not followed by a landfall
  #S - Change of status of the system
  #G - Genesis
  #T - Provides additional detail on the track (position) of the cyclone
  status <- factor(t[,4]) # storm status.
  # TD - Tropical cyclone of tropical depression intensity (< 34 knots)

```

```

# TS - Tropical cyclone of tropical storm intensity (34-63 knots)
# HU - Tropical cyclone of hurricane intensity (> 64 knots)
# EX - Extratropical cyclone (of any intensity)
# SD - Subtropical cyclone of subtropical depression intensity < 34 kt
# SS - Subtropical cyclone of subtropical storm intensity (> 34 knots)
# LO - A low that is neither a tropical cyclone, a subtropical cyclone,
# WV - Tropical Wave (of any intensity)
# DB - Disturbance (of any intensity)
#
# calculate latitude and longitude as signed floating point

lat_str <- t[,5]
n <- nchar(lat_str)
hemisphere <- substr(t[,5],n,n)
lat_str <- substr(lat_str,1, n -1) # N or S suffix denote hemisphere
lat <- as.numeric(lat_str)

lat <- ifelse( hemisphere == 'S', lat * -1., lat)
#
# calculate longitude as floating point.
#
lon_str <- t[,6]
n <- nchar(lon_str) # has W or E suffix.
lon_str <- substr(lon_str,1,n - 1)
lon <- as.numeric(lon_str)
hemisphere <- substr(t[,6],n,n)
lon <- ifelse( hemisphere == 'W', lon * -1., lon)

# decode the maximum wind observation.

max_wind <- as.integer(t[,7])
# values flagged as -99 are missing. So set to NaN.
# Don't use NA -- I get errors from cut()
# -99 flags missing values
max_wind <- ifelse(max_wind == -99, NaN, max_wind)

# use max wind to classify the storm at this point in time.
# The break values of max wind speeds is used to determine the
# the storm classification, per glossary definitions at nhc.noaa.gov
# TD - Tropical Depression, TS - Tropical Storm, and then
# hurricanes, category 1 through 5.

Cat <- cut(max_wind,
           breaks=c(1, 34, 64, 83, 96, 113, 137, 200),
           labels=c('TD', 'TS', 'Cat 1', 'Cat 2', 'Cat 3', 'Cat 4', 'Cat 5'))

# decode the minimum press
min_press <- as.array(as.integer(t[,8]))

```

```

# replace -999 pressure values with NA
min_press <- ifelse(min_press == -999 , NA, min_press)

# create new data frame of all track data.
df <- data.frame(uid, lrec, stormid, storm_year, name, rec_type,
  status, dtg,
  year, month, day, hour,
  minute, lat, lon, max_wind, min_press, Cat)

return(df) # return the dataframe
}

#
# readFile - read in a file of Atlantic storm track data.
# input: filename - path to the source data
# output: returns a dataframe of all track data in the file
#
readFile <- function(filename) {
  con <- file(filename, open="r")
  tryCatch(
  {
    allstorms <- readStorm(con) # read first storm
    nstorms = 1
    while( 1)
    {
      df <- readStorm(con)
      allstorms <- rbind(allstorms, df) # append
      nstorms = nstorms + 1
    }
  }, error = function(e) {
    print(e)
  }, finally = {
    close(con) # close the input file.
    print(sprintf("Read in %d storms", nstorms))
  }) # end of trycatch
  return(allstorms)
}

#
# saveHurtrack - read in the hurricane track data from 2013,
# and write the dataframe to a file
# input: none
# output: writes file hurtrack.Rda
#
saveHurtrack <- function() {
  df <- readFile("~/hurricanes/hurdat2-1851-2013-052714.txt")
  saveRDS(df, file="~/hurricanes/hurtrack.Rda")
}

```

```

}
# setup -- loads libraries needed for this project
# input: none
# output: none
#
setup <- function()
{
#
# load libraries needed for this project plotting histograms and maps of
# hurricane tracks
#
library(rworldmap) # map database of the world
library(scales) # has date_format() for labeling dates on plots
library(ggplot2) # for making plots
library(ggmap) # for plotting world maps
library(sp) # for spatial data
library(lubridate) # for the year() method
library(mapproj) # for map.grid() method
low_map <- getMap() # initialize a world map.
# read in the track data and save to a data frame
if ( file.exists('~/.hurricanes/hurtrack.Rda'))
{
  print ("The track data have been parsed already")
}
else
{
  print ("Parsing and saving track data now...")
  saveHurtrack()
}
}
#
# execute setup
#
setup()

```

```

## Loading required package: sp
## ### Welcome to rworldmap ###
## For a short introduction type :  vignette('rworldmap')
## Loading required package: maps

```

```

## [1] "The track data have been parsed already"

```

```

# load all the track data.
tr <- readRDS("~/hurricanes/hurtrack.Rda")
length(unique(tr$suid)) # returns 1771 as the number of storms

```

```

## [1] 1771

```

# Analysis

## Plotting all Tracks

One of the first classic plots of hurricanes is to plot the tracks of all the storms in the record on a map. Making a map plot required a lot of coding, including a lot of debugging and testing. But having made a plot routine I can plot any subset of the track data that I like. For starters, lets look at all tracks in history, and see where they started on the map.

```

#
# plot all the tracks in the dataframe
# input: df - dataframe of storms tracks to be plotted
# input: main_title = a main title label for the plot
# output: plot of all tracks with red dot at the starting point.
plotAllTracks <- function(df, main_title)
{
  #
  # set the window domain in latitude and longitude for the plot
  #
  minlat <- 5.
  maxlat <- 60.
  minlon <- -100.
  maxlon <- 0.
  # start a new plot
  plot(low_map,xlim=c(minlon,maxlon), ylim=c(minlat,maxlat))
  # plot the grid, but suppress the labels on grids -- too hard to read.
  map.grid(c(minlon,maxlon,minlat,maxlat),col='black', labels=FALSE)

  #
  # get a list of all the storms in the dataframe, identified by uid
  #
  storms <- unique(df$uid)
  #
  # plot the track for each storm,
  # and put a red circle at the starting point of each track
  #
  xlist <- array() # will store a list of x coordinates points to plot
  ylist <- array() # list of y coordinates to plot starting points
  for (storm in storms) # repeat for each storm
  {
    # select the storm by storm uid. need only lat and lon
    track <- subset(df, uid == storm,select=c('lat','lon'))
    # plot the lines.
    lines(x=track$lon, y=track$lat, col='blue')
    # use head to get the first record, the initial point, of each track,
    # to plot a solid (pch=16) red circle.
    ip <- head(track,1) # initial point
    # add the initial point to a vector of locations to plot a dot.

    xlist <- append(xlist, ip$lon)
    ylist <- append(ylist, ip$lat)

  }
  # plot the initial points of all tracks.
  points(xlist, ylist, col='red', pch=16, cex=0.3)
  # re-plot the base map so coastlines show up over the data.
  # add=TRUE prevents erasing the plot that is already there
  plot(low_map, xlim=c(minlon,maxlon), ylim=c(minlat,maxlat), add=TRUE)
  # add the labels

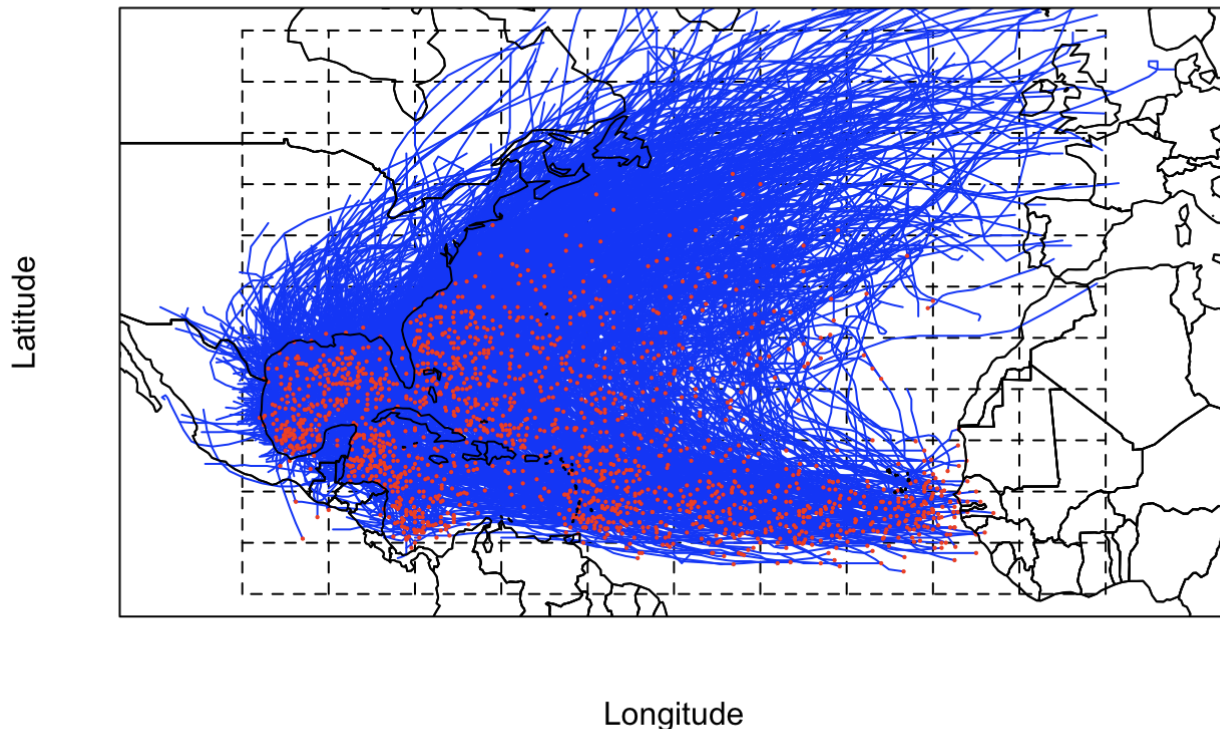
```

```

title(main=main_title, line=2, xlab='Longitude', ylab='Latitude')
box() # plot box around the graph.
}
#
# execute: get a base map, and plot all the tracks.
#
low_map <- getMap() # initialize a world map.
plotAllTracks(tr,'All Storm Tracks')

```

## All Storm Tracks



This map shows in blue lines all of the storms in the database, and I plotted a red dot at the start of each line so you can see where the storm originated. Books have been written about this hurricane data, but I will summarize key points.

The tracks are centered along the south to west, to northern sections of the Atlantic Ocean. The reason there are no (or very few) hurricanes along the eastern boundary of the Atlantic ocean is because the temperature of the water is too low. Generally hurricanes require water temperatures about 27 deg C or higher. The flow of air over the tropical Atlantic is dominated by the Bermuda High – a large cell of dry weather that rotates clockwise, and tends to push storms westward, then northward.

Hurricanes cannot form over land, as they derive their energy from evaporation of warm water from the ocean surface. So the red dots are all over water – with the exception that some storms may have formed over a small island in the Caribbean, or at least got a start there.

Also hurricanes do not form closer to the equator than about 5 degrees latitude. The rotation of the earth around the vertical axis approaches zero at the equator. Hurricanes require some rotation from the earth to help initiate the formation of a storm. Equator-ward of 5 degrees latitude there is insufficient rotation. So tropical cyclonic storms do not form near the equator.



Storms also do not form much in the northern half of the basin. This is because the water temperatures north of the Gulf Stream are too low to support hurricane formation. But once a vigorous storm gets going, the storm can last weeks as it travels northward, then dies out over the northern regions of cold water.

The diagram also indicates that there are preferred regions where lots of storms form. Hot spots are the western Caribbean Sea, the Gulf of Mexico, and the Bahamas.

From a data analysis perspective, this is a rather lousy plot to examine. Too much data is presented, and the lines saturate the plot. Yet you will see this diagram in textbooks. Later on I will present a heat map display of this data that provides more insight on storm activity.

## Distribution of Wind Speeds

The following plot shows the number of observations of wind speeds, color coded according to the storm classification. Storms are classified as Tropical Depression (TD), Tropical Storm (TS), or Hurricane (Category 1 through 5 – 5 being the most intense).

```
library(ggplot2)

#
# windhist - plot histogram of maximum wind speed -- all measurements made
# shows that the stronger the winds are the less frequently they occur.
#
windhist <- function()
{
  #
  # select the max_wind from each record, but omit records with NA values.
  #
  subdata <- subset(tr, select=c('max_wind', 'Cat'))
  # create a field called Cat, that is the storm category
  # for each measurement.

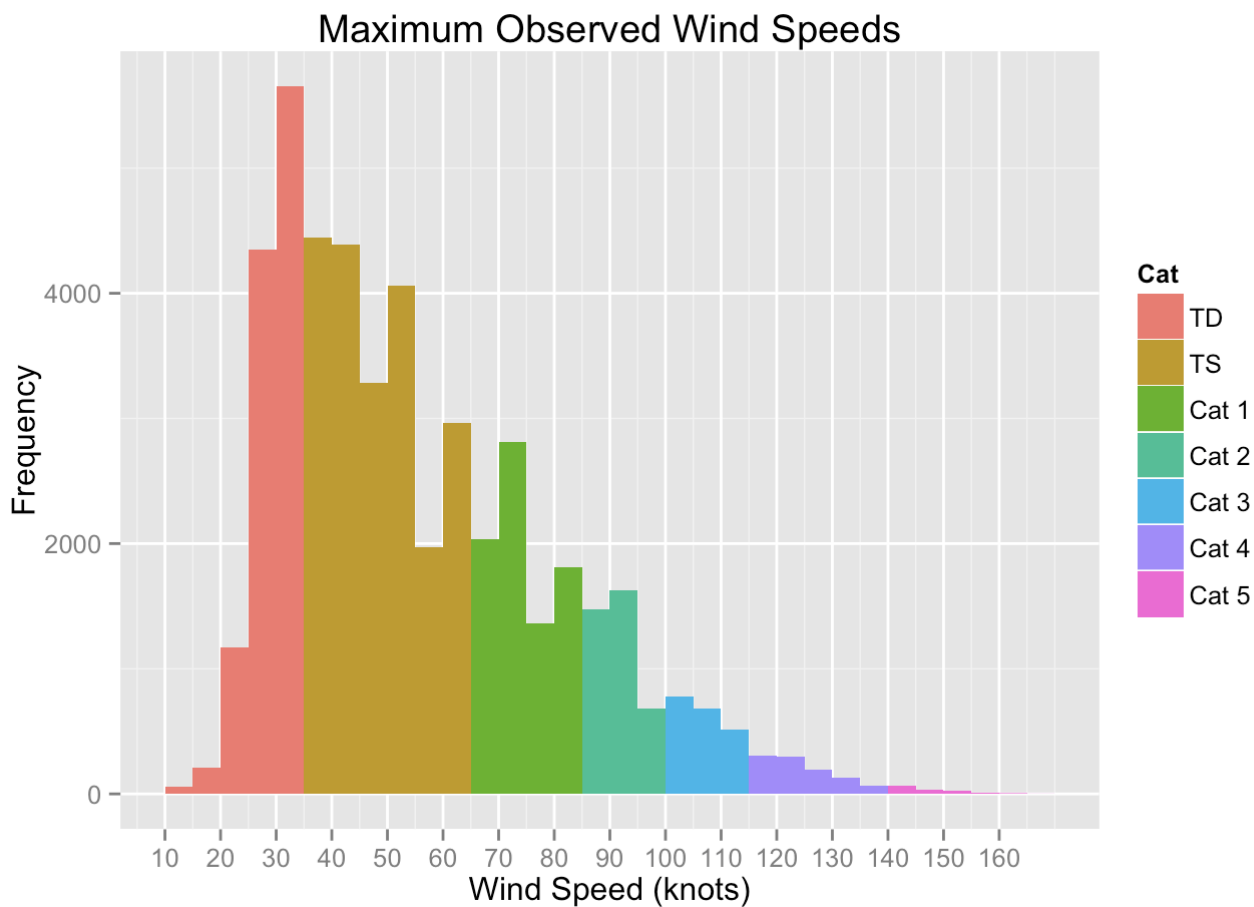
  ggplot(subdata) +
    geom_histogram(aes(max_wind, fill=Cat), binwidth=5) +

    scale_x_continuous(limits=c(10,170), breaks=seq(10, 160,10)) +

    labs(x='Wind Speed (knots)', y='Frequency',
         title='Maximum Observed Wind Speeds')

}

#
# execute windhist
#
windhist()
```



This plot shows very clearly how the more intense a storm is the less often it is observed. Category 5 storms in particular are extremely rare. And many of those stay at sea and are never an issue, or weaken before they hit land. In fact there are only a few Cat 5 hurricanes that have hit the United States. We can quickly make a report of all Cat 5 storms that made landfall in the basin using this query:

```
unique(subset(tr, Cat == 'Cat 5' & rec_type == 'L' ,
              select=c('name', 'year')))
```

```
##      name year
## 17943 UNNAMED 1935
## 28222 CAMILLE 1969
## 35782 GILBERT 1988
## 37300  ANDREW 1992
## 44466   DEAN  2007
## 44514  FELIX  2007
```

The unnamed hurricane of 1935 went through the Florida Keys and caused massive devastation, including knocking a train off the tracks, and killing all aboard. Andrew of 1992 was a category 5 when it went through Dade county Florida. Felix and Dean went into Central America. Camille devastated the Gulf coast.

Here is a track plot of all the landfalling Cat 5 hurricanes:

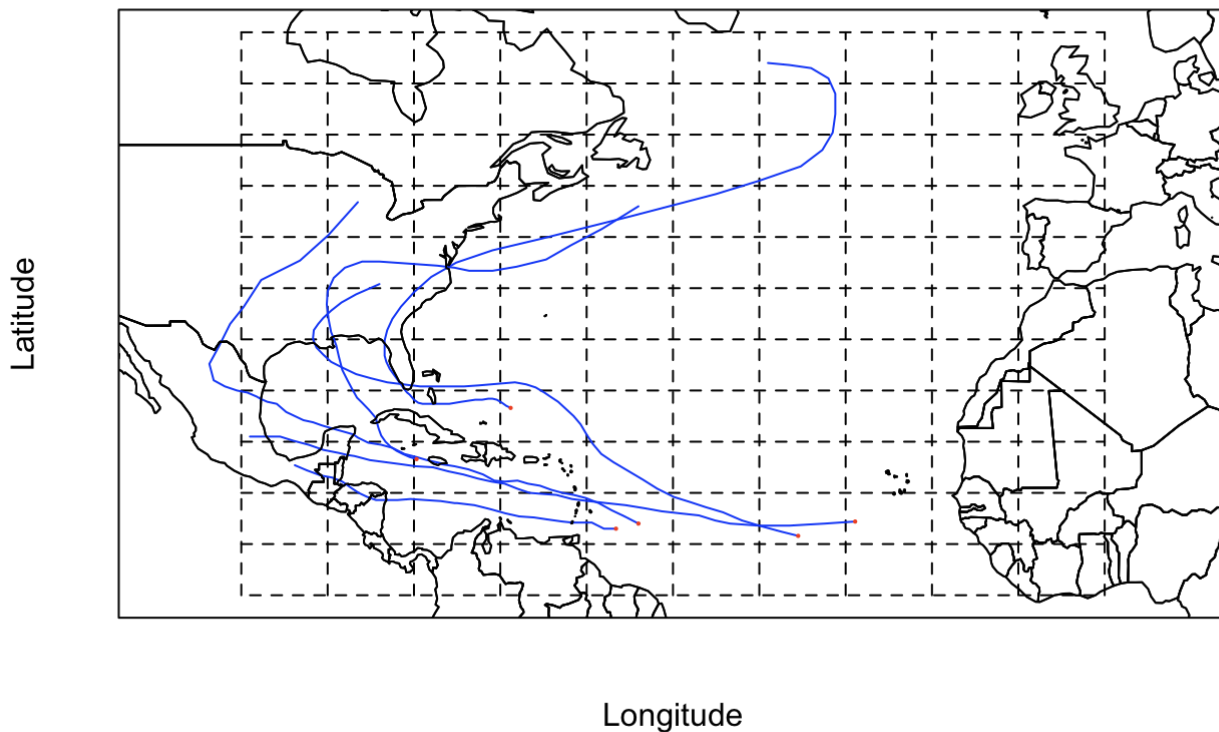
```

# select the primary key for each landfalling cat 5 storm
# rec_type has value 'L' to denote landfall.
# and the Cat variable is 'Cat 5'
landfall_storm_ids <- unique(subset(tr, rec_type == 'L' & Cat == 'Cat 5',
                                   select='uid'))

#
# now select all the track records where the storm is in that list of ids
#
tracks <- subset(tr, uid %in% landfall_storm_ids$uid)
#
# now plot the tracks of the landfalling cat 5 storms
#
plotAllTracks(tracks, 'Cat 5 at Landfall')

```

## Cat 5 at Landfall



This plot shows that the Cat 5 storms that make landfall tend to form far out at sea. It is a general thought among tropical meteorologists that the big intense killer storms get organized and develop out over the eastern Atlantic, where they have lots of time and warm water, with no land to get in their way, as they get organized and develop into intense storms.

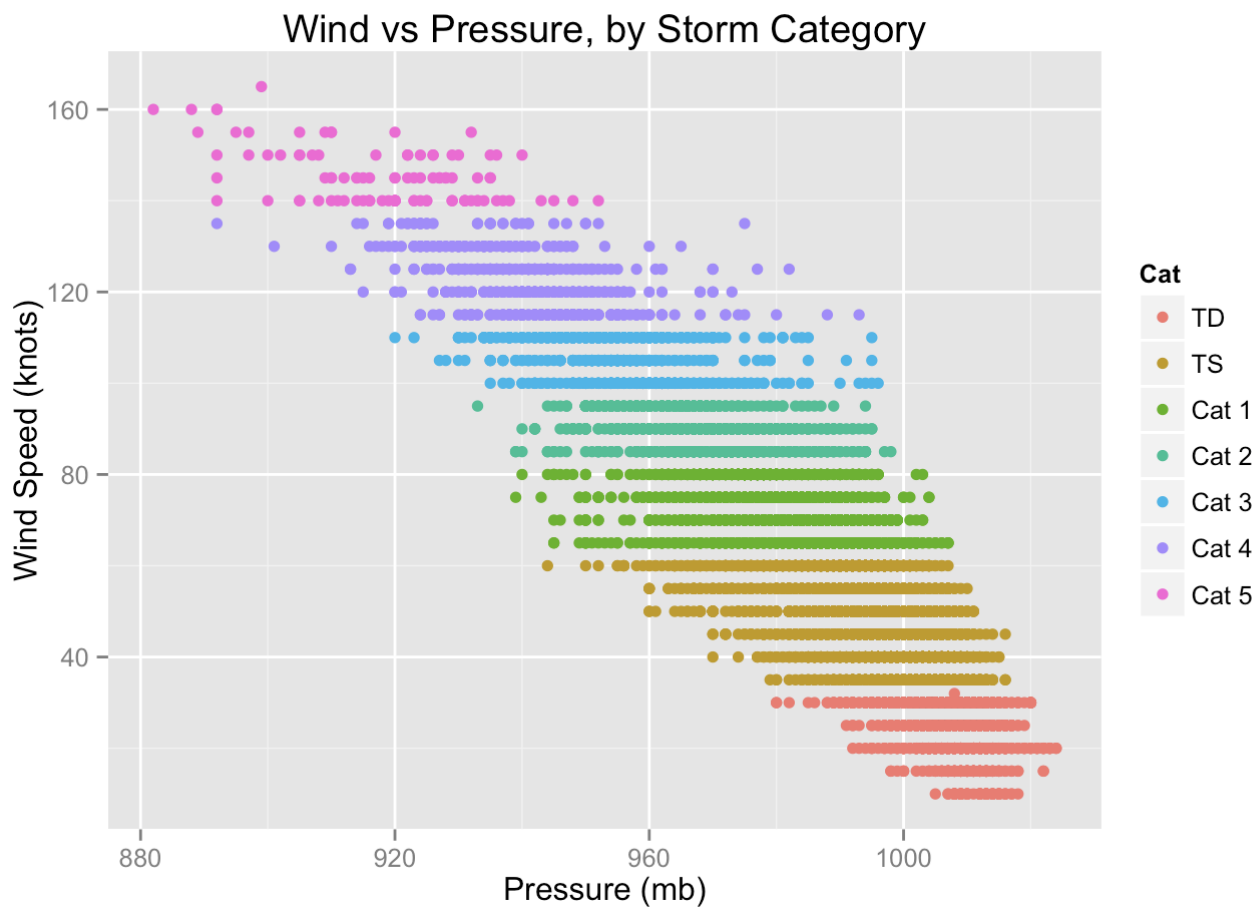
## Relationship between pressure and wind

The basic science of a hurricane is this: thunderstorms over the ocean, in the presence of usually some pre-existing cyclonic (i.e. rotating) disturbance in the lower atmosphere, combined with high sea surface temperatures, enables the storms to heat the atmosphere, and cause the air pressure to drop hydrostatically. Visualize that as a column of air warms, the column expands, the air becomes less dense, which means there is less mass in the column, which means that the air pressure at the surface (which responds to the mass of air above it) becomes lower. As the

pressure gets lower, air gets driven into the low pressure system from the environment where the pressure is higher. The first basic notion is that wind is the movement of air from higher to lower pressure. So air flows into the developing storm at low levels of the atmosphere (the first kilometer typically). The faster the air flows, the more moisture gets evaporated from the ocean, the more rain is produced by the rising clouds, and more heat gets released by condensation, the more the column of air expands and the pressure drops. This process results, in the case of a mature hurricane, in a positive feedback mechanism that leads to intensification.

As a result, low pressure is associated with high wind speeds. To visualize this, look at this scattergram of minimum pressure and maximum wind speed, colored by storm category:

```
#  
# plot wind vs press  
# input: dataframe of track records  
# output: scatterplot of max wind versus min pressure,  
# colored by storm category  
windpress <- function(df)  
{  
  
  ggplot(df) +  
    geom_point(aes(min_press, max_wind, color=Cat,name='Category'),  
              alpha=1.0) +  
    labs(x='Pressure (mb)', y = 'Wind Speed (knots)',  
         title='Wind vs Pressure, by Storm Category')  
  
}  
windpress(tr) # plot the scattergram of wind vs pressure  
  
## Warning: Removed 30387 rows containing missing values (geom_point).
```



You can see that there is roughly an inverse linear relationship between pressure and wind. This relationship is employed by hurricane forecasters and analysts. Just by having a single pressure reading, say from a ship in the middle of a storm they can estimate what the maximum winds of the storm likely are, or soon will be.

An important thing to know about the damage done by wind is that the damage increases with the square of the wind speed. It is the kinetic energy of wind, or equivalently the wind pressure per square area, that damages structures. Generally speaking well built houses can stand up to a Cat 1 or 2 storm. But Cat 3, 4, and 5 will damage or destroy even well built and storm-protected houses.

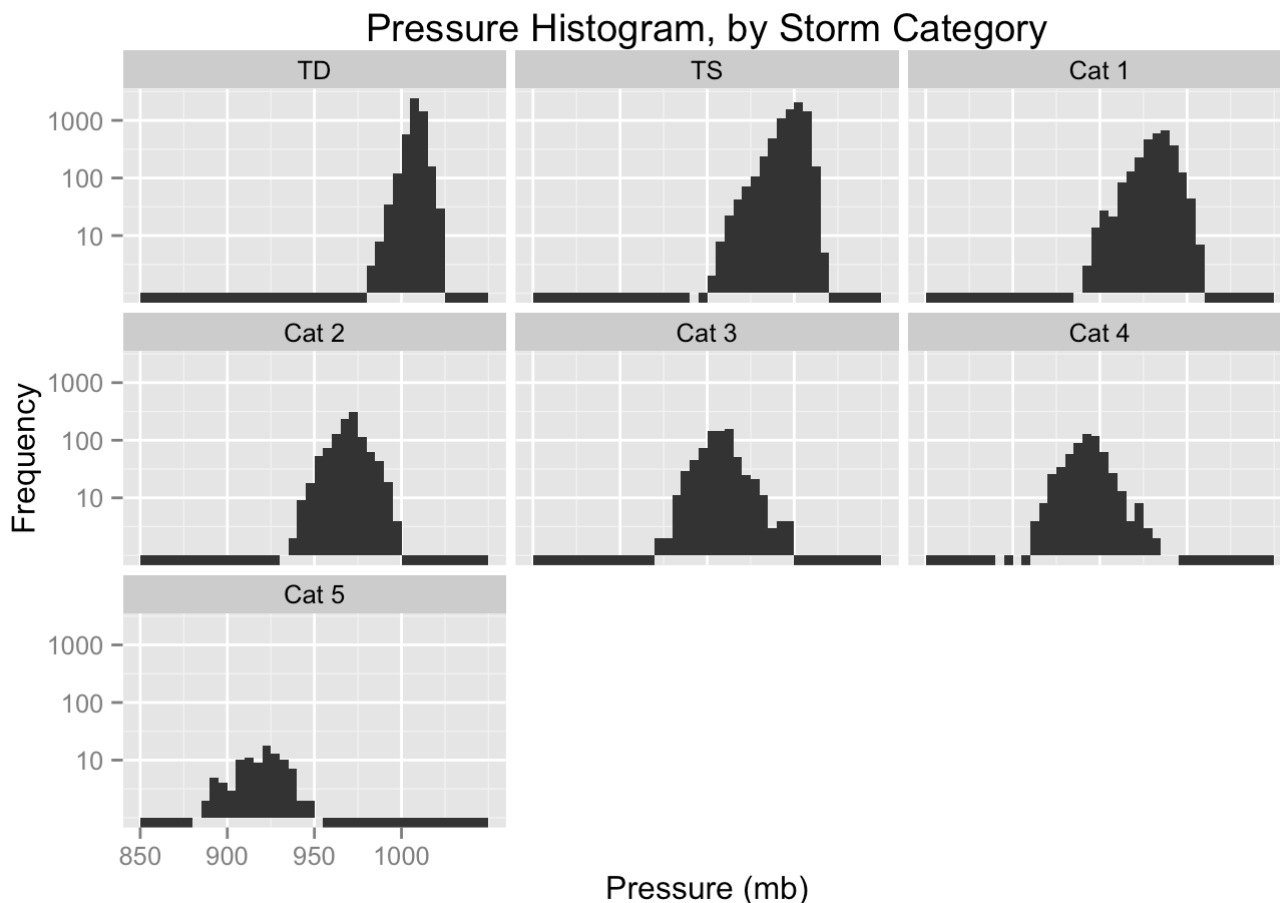
As Cat 5 winds are rare, so too are low pressures associated with Cat 5 storms. To illustrate, here is a histogram of pressures, facet wrapped by storm category:

```

# plot histogram of pressures, faceted by storm category at the
# time of observation
# This shows how intense storms have low pressure,
# and low frequency of occurrence.
# input: df - dataframe of track data
# output: plot of histogram
library(ggplot2)
pressHist <- function(df)
{
  # select just the max wind and min press, and Cat fields
  subdata <- na.omit(subset(df, select=c('max_wind', 'min_press', 'Cat')))

  # now plot the histogram.
  # position="identity" eliminates warning messages about stacking.
  # but otherwise has no effect.
  ggplot(subdata) + geom_histogram(aes(min_press), binwidth=5,
                                   position="identity") +
    scale_y_log10( breaks=c(10,100,1000)) +
    scale_x_continuous(limits=c(850,1050), breaks=seq(850, 1000,50)) +
    facet_wrap(~Cat) +
    labs(x='Pressure (mb)', y = 'Frequency',
         title='Pressure Histogram, by Storm Category')
}
pressHist(tr) # plot the pressure histogram

```



Cat 5 pressures – roughly below 950 mb, have been observed order of tens of times, while pressures associated weak storms have been observed thousands of times.

# Storm Frequency During the Year

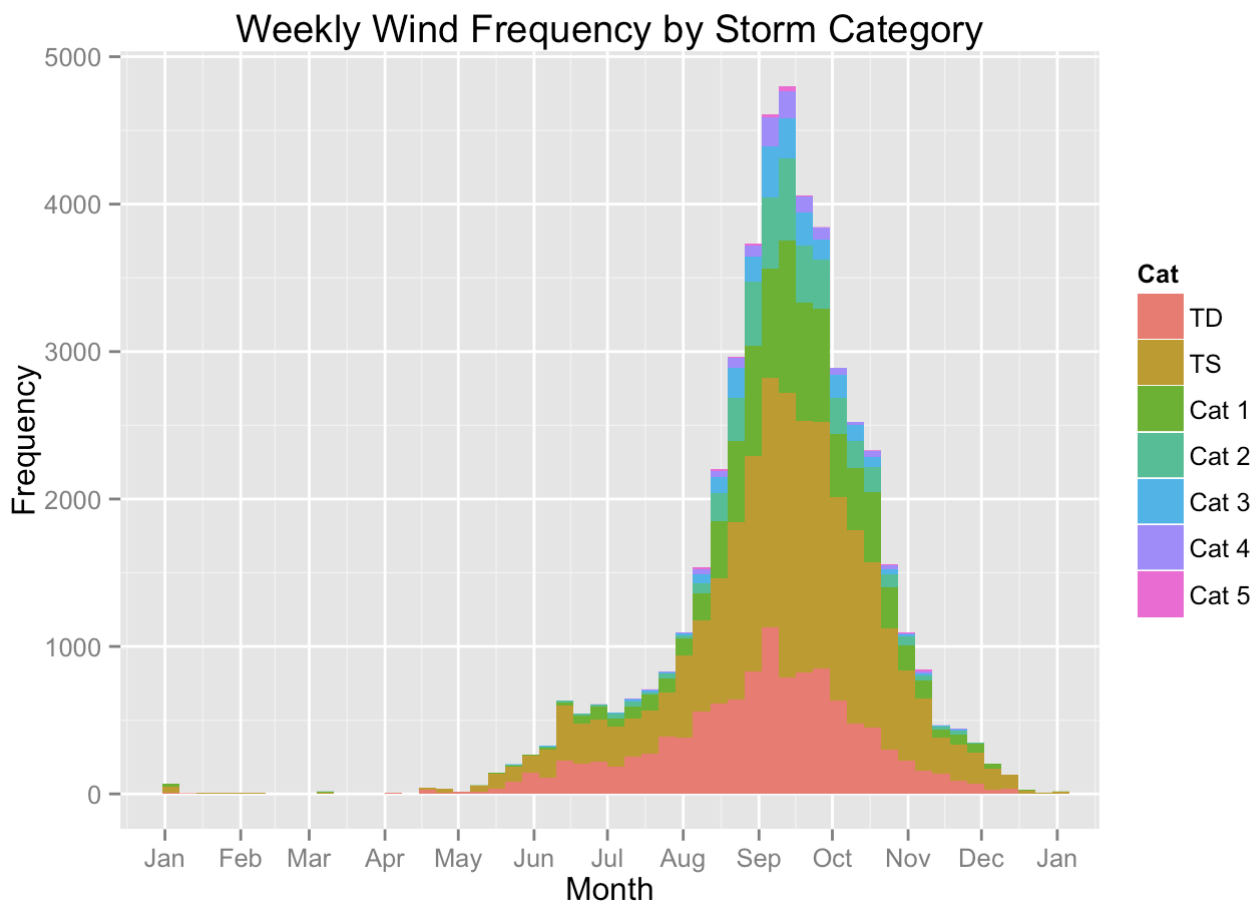
The Atlantic hurricane season runs officially from June 1 through November 30 of each year. The vast majority of storms form during this period. The next plot shows the frequency of occurrence of storm wind speeds for each week of the year. The wind speeds are colored according to the storm category.

```
#
# plotdoyHist - plot the storm frequency by day of year
# input: dataframe of tracks
# output: plot of storm frequency by day of year
#
plotdoyHist <- function(tracks) {
  # select just the date and the category
  first <- subset(tracks,select=c('dtg', 'Cat'))
  # set all the years constant, for an intrayear histogram.
  year(first$dtg) <- 2010

  # scale_x_date plots out the months of the year, from Jan through Dec.
  # binwidth = 7 makes this a 7-day (1 week) totals histogram.
  # Daily totals are too noisy.

  first$date <- as.Date(first$dtg)
  first <- na.omit(first) # filter out any data where the category is NA
  #
  # plot histogram of storms by week of the year, colored by storm category
  ggplot(first) +
    geom_histogram(aes(date, fill=Cat), binwidth=7) +

    scale_x_date(limits =
                  as.Date(c('2010-01-01','2010-12-31')),
                  labels = date_format("%b"), breaks='1 month') +
    labs(x='Month', y='Frequency',
         title='Weekly Wind Frequency by Storm Category')
}
plotdoyHist(tr) # plot the histogram
```



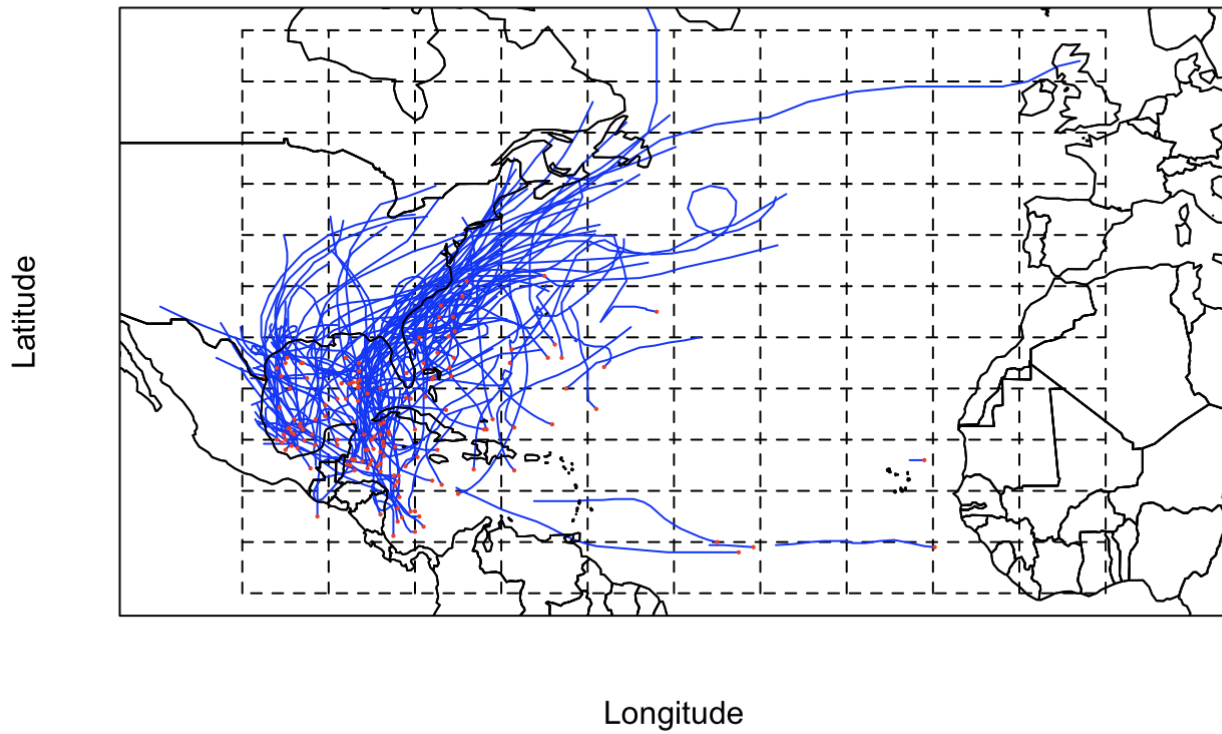
This graph shows that the peak of the season is the first week of September. Interestingly the early storms of May and June tend to be weak, and the strong hurricanes tend to form later in summer. Meteorologists explain this phenomenon by noting that the early storms tend to form close to the continent, often forming along disturbances created by cold fronts that stall and die in the Gulf of Mexico.

To illustrate this, let's look at the tracks of all storms that formed in May and June – the early part of the season.

```
mayjune <- subset(tr, month >= 5 & month <= 6)
plotAllTracks(mayjune, 'All May and June Storms')
```



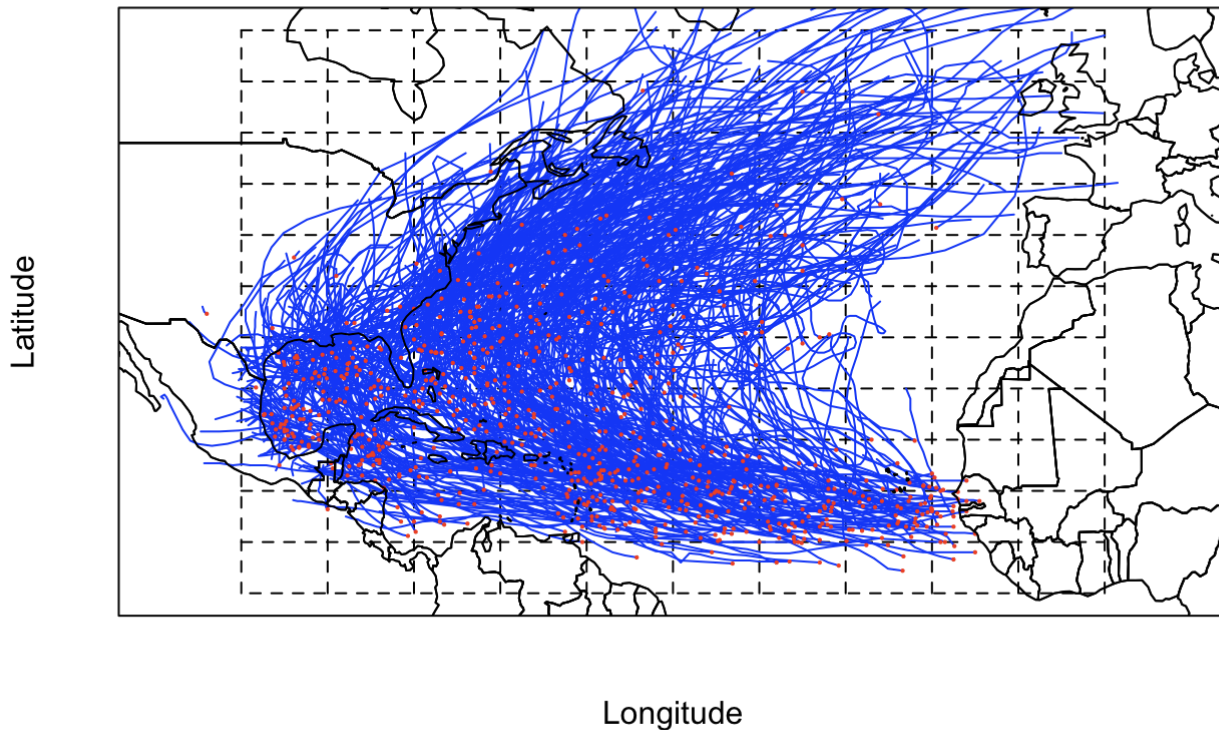
## All May and June Storms



Now contrast that with storms that form in September:

```
sep <- subset(tr, month==9)
plotAllTracks(sep, 'All September Storms')
```

## All September Storms

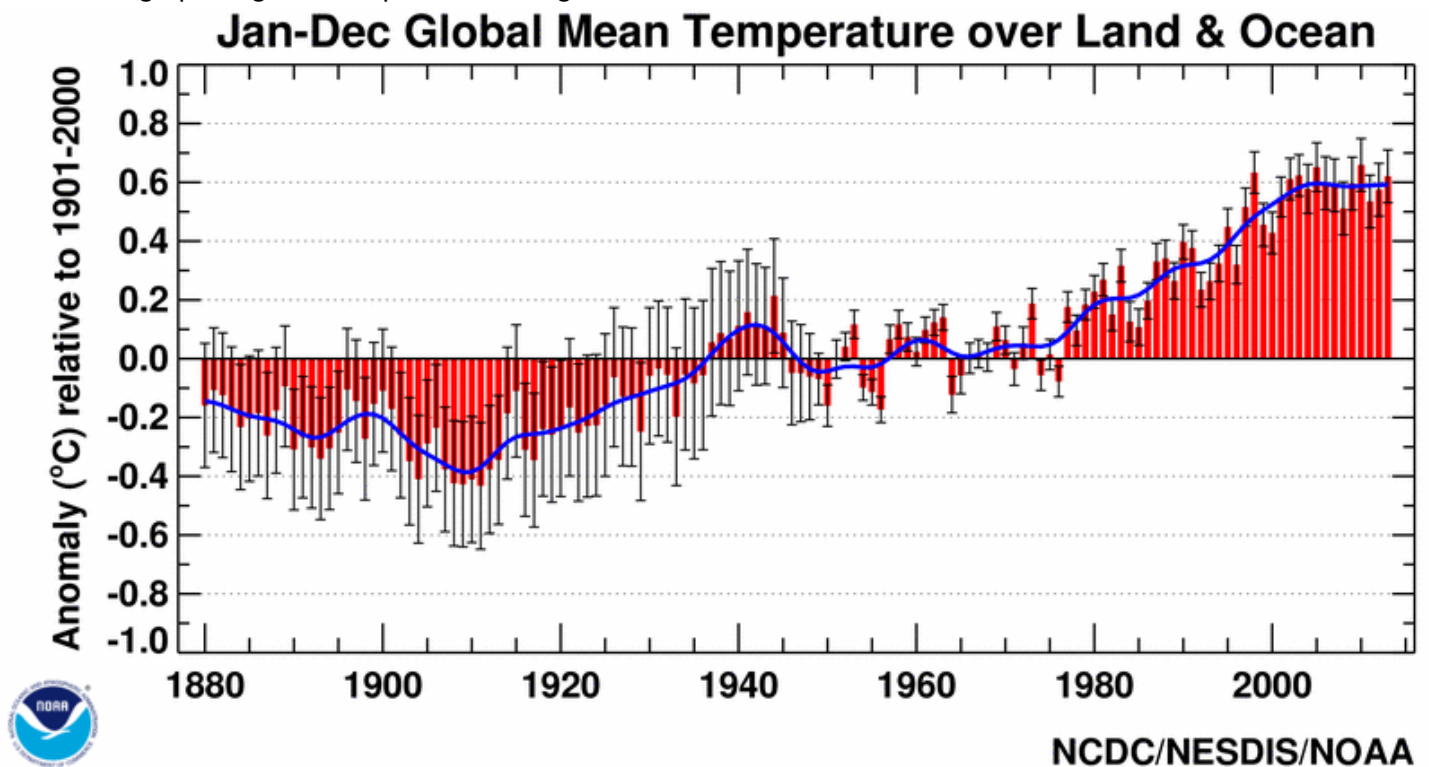


You can see that the early season storms form close in to the United States. And that the late season storms form all over the place, from the far eastern Atlantic to the western Caribbean.

## Climate Change and Tropical Cyclonic Storm Activity

A current topic of discussion about hurricanes is whether or not hurricane activity is affected by the increasing global temperatures of the earth. One theory is that a warmer planet will produce warmer oceans, which will produce more hurricanes, or at least hurricanes that are more intense than before.

The classic graph of global temperature change since the 1800's looks like this:



Ref: <https://www2.ucar.edu/sites/default/files/news/2014/201301-201312.png>  
(<https://www2.ucar.edu/sites/default/files/news/2014/201301-201312.png>)

This graph shows the change in the globally averaged surface temperature had a minimum temperature around 1920, followed by a warming trend that continued until about 1940. Then temperatures were fairly steady until 1970, following by a steady rise till about 2010, and a plateau since then.

Alternate theory suggests that sea surface temperature has never been a limiting factor for hurricanes. The Atlantic is always warm enough every year to support lots of hurricanes. But some years are very active and others are very quiet. Other meteorological factors, such as the amount of wind shear over the Atlantic, or the prevalence or absence of initial disturbances, controls the storm activity.

We can start looking at this question first by looking at the history of storm activity from 1851 to 2013. The next figure shows the frequency of occurrence of wind speeds on an annual basis, colored by storm category.

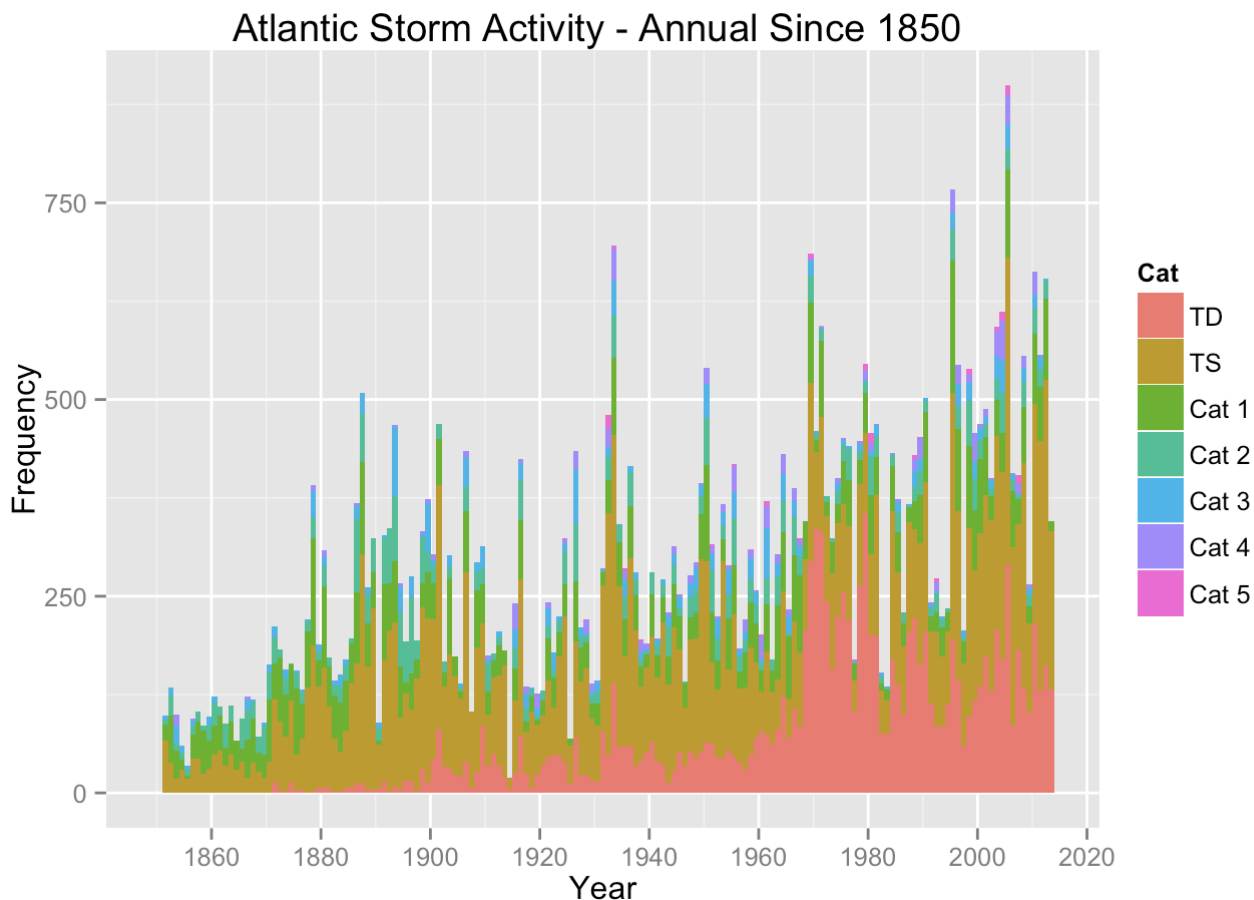
```

# plotyearHist -- plot histogram of storm frequency since 1850
# input: dataframe of all tracks
# output: plot of storm frequency.
#
plotyearHist <- function(df) {

  # get just the date and category for each record
  f <- subset(df, select=c(dtg,Cat))
  f <- na.omit(f) # filter out any data where the category is NA

  # prepare a vector of Date types -- which is required by ggplot.
  # binwidth = 365 means sum over a whole year.
  #
  dates <- as.Date(f$dtg)
  ggplot(f ) + geom_histogram(aes(x = dates, fill=Cat), binwidth=365) +
  scale_x_date(breaks = '20 years', labels = date_format("%Y") ,
               limits = as.Date(c('1849-01-01','2014-01-01'))) +
  labs(x='Year', y='Frequency',
       title='Atlantic Storm Activity - Annual Since 1850 ')
}
plotyearHist(tr) # plot the time series histogram

```



There is a lot of interesting information here. The variance from year to year seems rather high: active years are often followed by quiet years. This inter-annual variability has been ascribed to the effects of El Nino. El Nino is a phenomenon in which the ocean currents change direction over the eastern Pacific Ocean once every couple years.

That change in ocean current changes ocean temperature, which changes the weather conditions over the ocean, which in turn increases the wind shear over the tropical Atlantic. Wind shear in turn tends to inhibit storm formation and disrupt storm intensification, as it disrupts the integrity of the warm column of air that is needed for tropical cyclonic storm formation.

Most startling about this graph is the clear upward trend in *recorded* activity since 1850. I emphasize that this data represents the observed and recorded storms. The number of storms observed increased dramatically beginning in 1968, which is when the National Hurricane Center had regular weather satellite imagery available to find storms at sea, especially when the storms were forming. Hence you see that by 1970 the number of tropical depression wind observations increased dramatically. This increase is attributed to better observational capabilities, not a real increase associated with the storm activity. There seems to be an increase in storm activity from 2000 and on, compared to earlier decades. But much of that increase seems to be associated with tropical depressions and tropical storms.

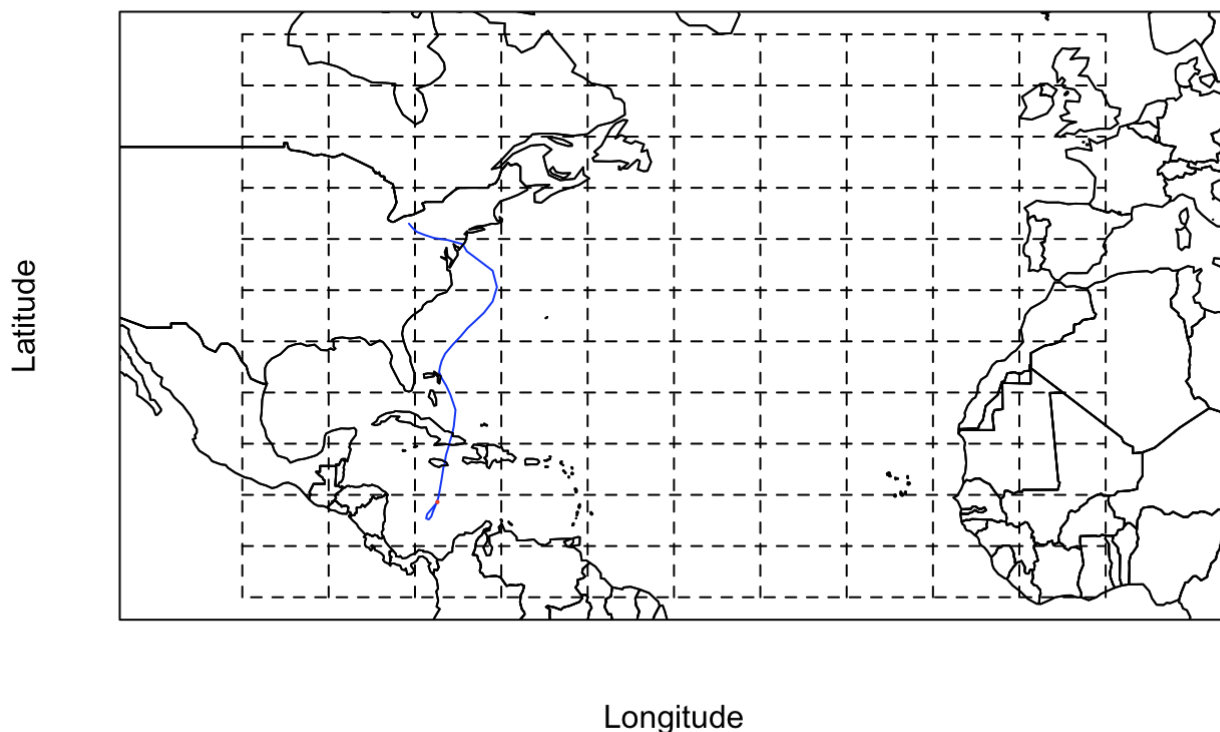
# Climate Change and Hurricane Risk

## Using a Heat Map

The storm track maps above are helpful when you want to look at just one storm, or a couple storms. For example, suppose you wanted to see the track for hurricane Sandy:

```
sandy <- subset(tr, name=='SANDY' & year==2012)
plotAllTracks(sandy, 'Superstorm Sandy')
```

### Superstorm Sandy



This graph shows that Sandy formed in the Caribbean Sea, and moved northward, missing most of the United States. But finally, and – unlike most other storms that head out to sea – got pushed back to the mainland by a fair weather system over the Canadian Maritime provinces. Bad luck for New York and New Jersey!

But I want to understand where hurricane activity is most pronounced, and find out what areas are most likely to experience hurricane force winds. The “spaghetti” plot (i.e the first figure with thousands of lines on it) I made earlier did not help much. So I created another plotting program where I create a grid of 1 deg latitude by 1 deg longitude cells, and count up the number of times a storm tracks over each grid. Then I plot a color coded heat map using that data.

Here is a heat map of all storm activity:

```

#
# plotHeatMap - plot track data converted to a grid of
# 1 deg lon by 1 deg lat
# input: track_data - data frame of storm tracks
# input: sub_title - a string describing what sample of data is being plotted
# output: plots heat map of the input data.
plotHeatMap <- function(track_data, sub_title)
{
  #
  # set the window domain in latitude and longitude for the plot
  #
  minlat <- 5.
  maxlat <- 60.
  minlon <- -120.
  maxlon <- 0.
  increment = 1.0 # 1 degree of latitude between sampling boxes
  lats <- seq(minlat, maxlat, increment) # latitudes to plot data within
  lons <- seq(minlon, maxlon, increment) # longitudes to plot data at
  # calculate the number of rows and columns in the grid needed
  # to match the domain
  xdim <- as.integer((maxlon - minlon)/increment)
  ydim <- as.integer((maxlat - minlat)/increment)
  # create an array to count tracks in. initialized to zero.
  counts <- array(0, dim=c(ydim,xdim))

  #
  # count up the number of times that tracks pass through a lat-lon box
  # for each track record, using the latitude and longitude of the
  # record to calculate row and column indices.
  # This is done as a vector operation.
  # ifelse() method to used to check the bounds;
  # any indices out of bounds get set to NA.
  # calculate row index
  rows <- as.integer((track_data$lat - minlat) / increment)
  rows <- ifelse(rows > 0 & rows <= ydim, rows, NA) # bounds check
  # calculate column index
  cols <- as.integer((track_data$lon - minlon) / increment)
  cols <- ifelse(cols > 0 & cols <= xdim, cols, NA) # bounds check

  #
  # for each point in list of columns and rows, count how many times
  # each location occurs
  # Note: there may be vectorized way to do this,
  # but I couldn't figure it out!
  #
  for( i in 1:length(cols))
  {
    counts[rows[i],cols[i]] <- counts[rows[i],cols[i]] + 1
  }
}

```

```

#
# make up a color table. This is a heat map type of scheme with 8 colors
# Places with higher hurricane frequency will have hotter color
#
color_table <- array(c('grey', 'purple','blue','green',
                      'yellow','orange','red', 'white'))

#
# create a table of color indices for each grid point.
# I rescale the data to range from 1 to 8
# Convert from counts to a color index
color_index <- (8. * counts /max(counts) + 1.)
# apply bounds checking
color_index <- ifelse(color_index >0 & color_index <=8,
                      color_index, NA)

# plot the base map.
plot(low_map, xlim=c(minlon,maxlon),ylim=c(minlat, maxlat))
#
# create 2 arrays to store the lat lon coordinates of a lat/lon box to plot
#

xx <- array(dim=4) # store longitude for ploygon plot
yy <- array(dim=4) #stores latitude for polygon plot
for(i in 1:xdim) # repeat for each column in grid
{
  # write x coordinates of the lat/lon box to color.
  xx[1] <- lons[i]
  xx[2] <- xx[1]+ increment
  xx[3] <- xx[2]
  xx[4] <- xx[1]

  for(j in 1:ydim) # repeat for each row in grid
  {

    # plot box where there is some data.

    if( counts[j,i] > 0)
    {
      #write the y coordinates of the box
      yy[1] <- lats[j]
      yy[2] <- yy[1]
      yy[3] <- yy[2] + increment
      yy[4] <- yy[3]
      # look up the color to plot from the color table.
      cc <- color_table[color_index[j,i]]
      # plot the filled colored box
      polygon(xx,yy, col=cc, border=cc)
    }
  } # end for j
} # end for i

```



```

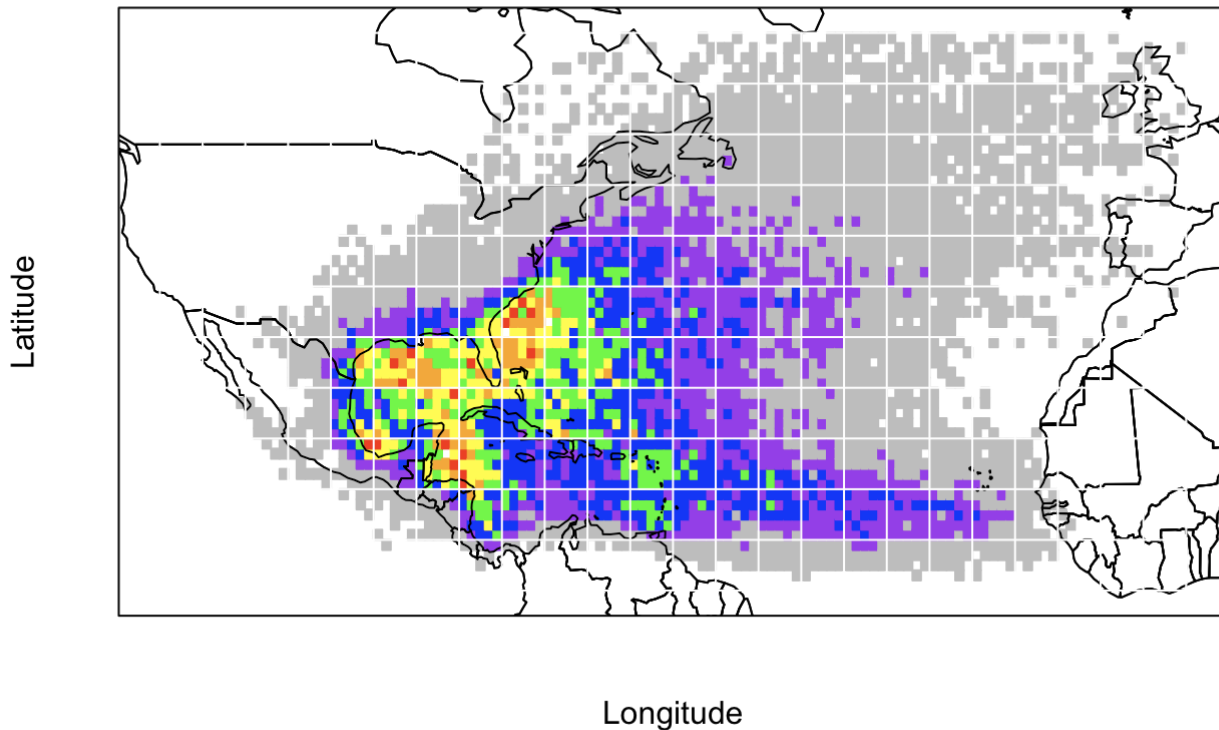
#
# plot the base map again, to overwrite any plotted data
# Use the same domain value set above.
# add=TRUE is needed so that the existing plot get updated, not replaced.
#

plot(low_map, xlim=c(minlon,maxlon),ylim=c(minlat, maxlat), add=TRUE)
#
# put a grid on top of this.
# Turn off labels, because they are too hard to read, and they cover up data.
#
map.grid(c(minlon,maxlon,minlat,maxlat),
        # plot grid lines every 5 x 5 degrees
        nx=(maxlon - minlon)/5., ny=(maxlat - minlat)/5. ,
        col='white',labels=FALSE, lty='solid')
# plot the title and sub title.
# line=1 gives good separation between titles and box.
title(main=paste('Storm Track Heat Map - ', sub_title),
      line=2, xlab='Longitude', ylab='Latitude')
box() # plot box around the graph.


} # end plotHeatMap()
#
# Execute plotting heat map of all storms
#
plotHeatMap(tr,'All Storms')

```

## Storm Track Heat Map - All Storms



This graphic shows in hot colors (yellow, orange, red, and white) the areas with relatively maximum storm activity, and cool colors (grey, purple, blue and green) have relatively less activity. This clearly shows the maximum activity near the coastline of the U.S., off the Florida and Carolina and Gulf coasts. A another hot spot is the southwest Caribbean Sea. Also the pattern extends out into the deep tropical Atlantic.

This graphic is useful for public awareness. It clearly shows the risk to all of Florida (which most seasoned Floridians already know; but new comers may not). The risk to cities along the Gulf Coast is likewise high. Hurricane Katrina taught that lesson.

But this graph shows all the storm activity – including low impact depressions and tropical storms. Even Cat 1 hurricanes can cause massive damage due to the storm surge, even though wind damage may be limited to downed trees and power lines. For example, Hurricane Sandy, at the time of landfall, had maximum winds of just 70 knots:

```
sandy_landfall <- subset(tr,name=='SANDY' & year == 2012 & rec_type == 'L' & day == 29)
sandy_landfall$max_wind
```

```
## [1] 70
```

Technically Sandy was already downgraded to a tropical storm at the time of landfall. But storm surge had built up for days, and the flooding caused massive damage. So lets look at the activity associated only with landfalling hurricanes.

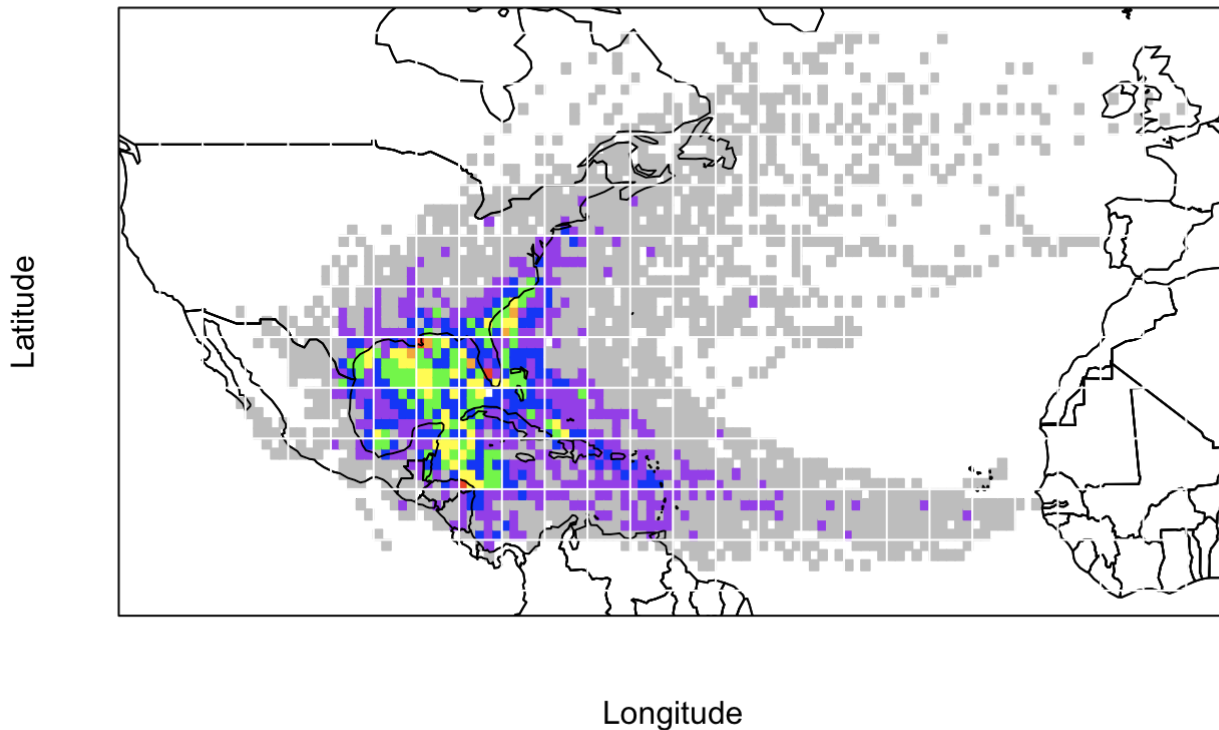
```

#
# landfall -- get tracks of landfalling hurricanes
# Uses this rule to identify landing falling hurricanes:
# any storm that has at least one record containing
# a rec_type equal to "L" (landfall),
# and that has a hurricane status, status == HU, at the same time.
#
# input: df - dataframe of all storm tracks
# return: tracks - dataframe subset containing only landfalling hurricanes
#
landfall <- function(df)
{
  # find the unique list of storm ids that match the rule.
  landfall_storm_ids <- unique(subset(df, rec_type == 'L' & status == 'HU',
                                     select='uid'))

  #
  # now select all the track records where the storm is in that list of ids
  #
  tracks <- subset(df, df$uid %in% landfall_storm_ids$uid)
  return(tracks)
}
#
# execute - get landfall tracks
lfh <- landfall(tr)
# plot heat map
plotHeatMap(lfh, 'Landfalling Hurricanes')

```

## Storm Track Heat Map - Landfalling Hurricanes



This plot is very interesting; it shows all the coastlines that are vulnerable to any kind of tropical cyclonic storm activity are vulnerable to hurricanes. There isn't any geographical place that just gets depressions or tropical storms. And the place of relatively highest risk looks to be about Miami.

## Final Plots and Summary

Finally the core question I want to answer is what areas are most at risk of devastating hurricanes? And is this risk increasing, decreasing, or staying about the same as the climate changes?

Let's try looking at a track map and a heat map of just major hurricanes – Cat 3, 4 and 5. Use the same algorithm as for all landfalling hurricanes, but select only those that are Cat 3, 4, or 5 at the time of landfall.

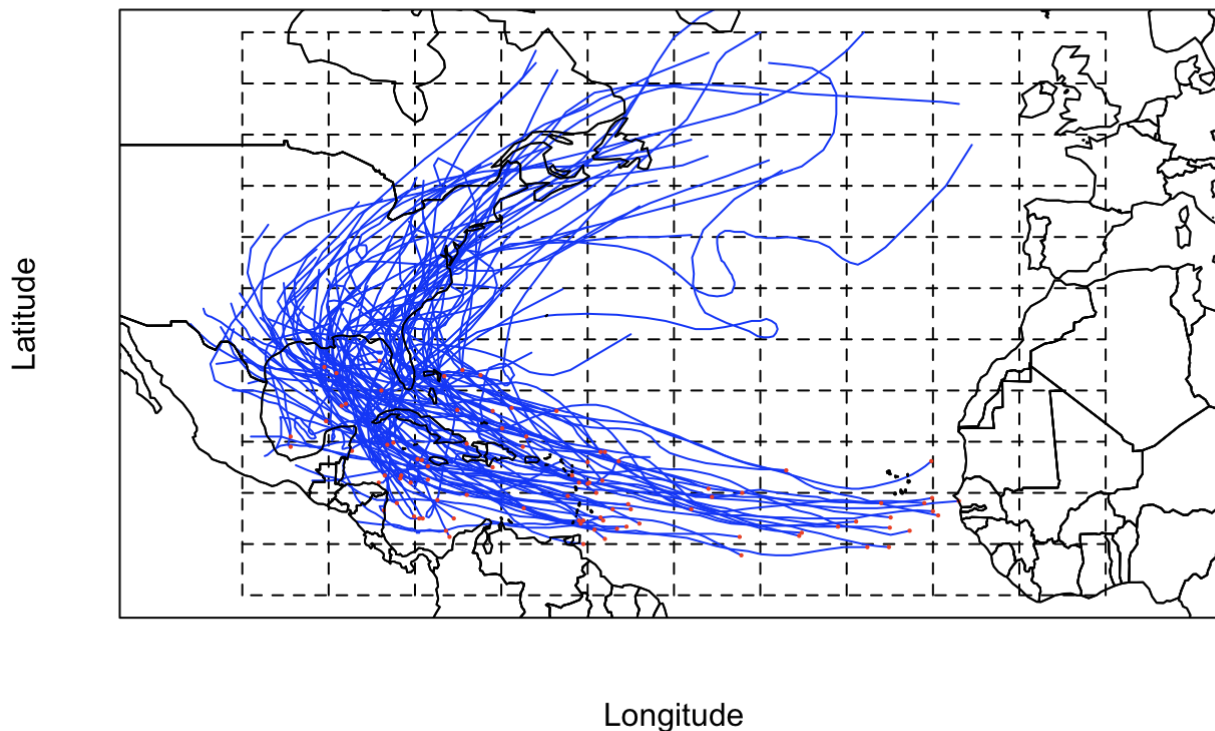
```

# major_landfall -- get tracks of landfalling hurricanes
# that are Cat 3, 4, or 5 at time of landfall.
# Uses this rule to identify landing falling hurricanes:
# any storm that has at least one record containing
# a rec_type equal to "L" (landfall),
# and that has a Cat value of Cat 3, Cat 4 or Cat 5 at the same time.
#
# input: df - dataframe of all storm tracks
# return: tracks - dataframe subset containing only landfalling hurricanes
#
major_landfall <- function(df)
{
  # find the unique list of storm ids that match the rule.
  major_hurr_ids <- unique(subset(df, rec_type == 'L' &
                                Cat %in% c('Cat 3', 'Cat 4', 'Cat 5'),
                                select='uid'))

  #
  # now select all the track records where the storm is in that list of ids
  #
  tracks <- subset(df, df$uid %in% major_hurr_ids$uid)
  return(tracks)
}
major <- major_landfall(tr)
# plot the track map
plotAllTracks(major, 'Major Hurricanes at Landfall')

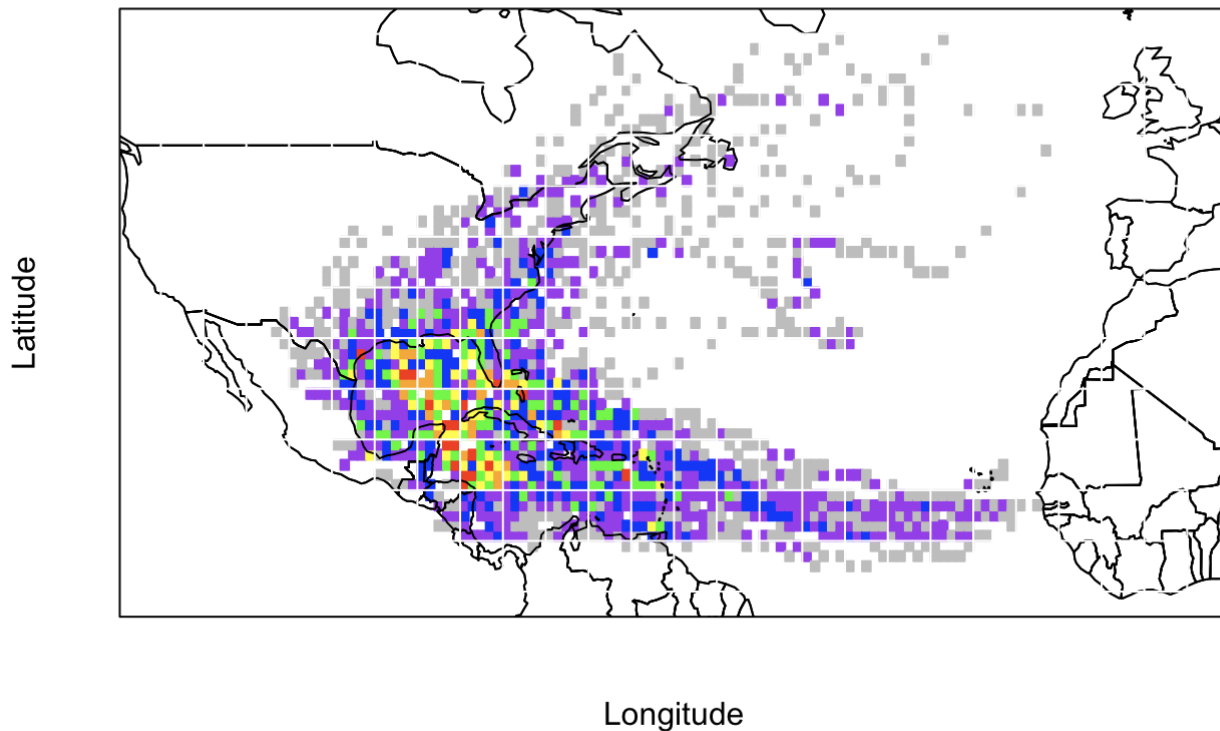
```

## Major Hurricanes at Landfall



```
# and the heat map
plotHeatMap(major, 'Major Hurricanes at Landfall')
```

## Storm Track Heat Map - Major Hurricanes at Landfall



The track map does not provide much insight into where the greatest risk of landfalling major hurricanes resides – largely because the lines cross so much. The heat map is more useful.

Two distinct lines of approach are evident: from South Florida and extending to the southeast is a preferred approach for major hurricanes. But even more active is the Gulf Coast, centered around New Orleans, and extending south east into the hot spot of activity in the Western Caribbean. This climatology of major hurricanes shows that New Orleans and surrounding gulf coast regions must never let their guard down regarding hurricane awareness and preparation.

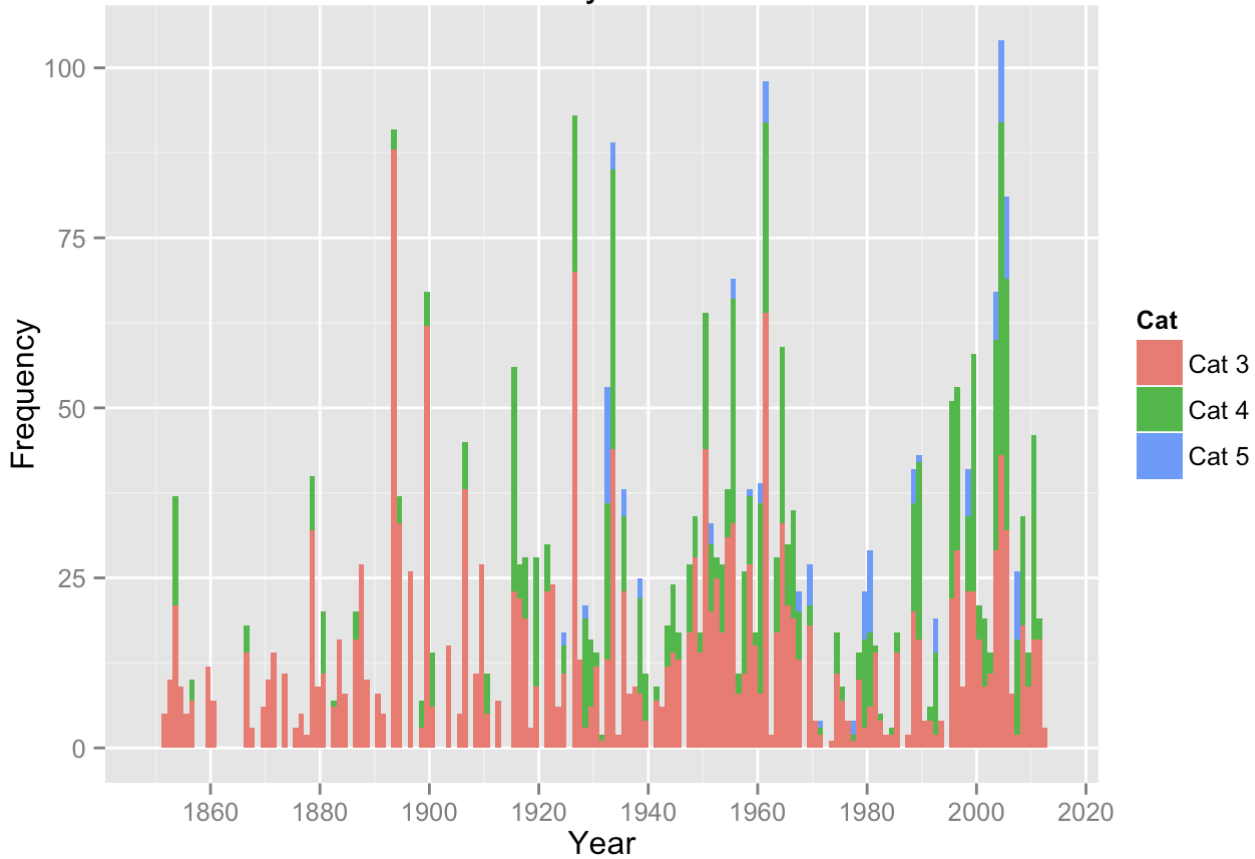
## Is Climate Change Affecting Major Hurricane Activity?

A key question being asked, by scientists, politicians, and media who cover this topic (including a world of bloggers either with or without scientific training) is this: Are major hurricanes becoming more common due to global warming? One way to address this question, for the Atlantic basin, is to look at the time histogram of major hurricanes – the Cat 3, 4 and 5 storms.

Category 3, 4 and 5 hurricanes do most of the damage, and cause most of the deaths, even though they are a relatively small share of the storms. There is a theory that once a hurricane gets organized – has the initial disturbance, grows over water without getting sheared apart by the environmental wind – that a warmer climate will make more hurricanes grow into intense hurricanes. Let's look at how intense hurricane activity fared:

```
# select records that are of category Cat 3,4 and 5
major_storms <- subset(tr, Cat %in% c('Cat 3', 'Cat 4', 'Cat 5'))
# and plot the timeline
plotyearHist(major_storms)
```

## Atlantic Storm Activity - Annual Since 1850



This suggests that the 1950s and 60s were very active with major storms, but this was a period of stable global temperature. Then the 1980s were quiet in terms of storms, while the climate was warming. And the 2000s were active again, while the climate was stable but very warm. So there is no obvious association between temperature trends and hurricane activity.

Another way to look for a climate signal is to look at the trend over time in the observed minimum pressure. Since pressure is lowest in intense hurricanes, we might see a trend associated with climate change.

```

#
# presshist - pressure history
# look at the time varying minimum pressure of major hurricanes
# looking for a climate signal.
# input: dataframe of all tracks
# output: time vs. minimum pressure plot
#
presshist <- function(df)
{
  # get just the date and category for each record
  f <- subset(df, select=c(dtg,min_press))
  # prepare a vector of Date types -- which is required by ggplot.

  dates <- as.Date(f$dtg)
  ggplot(f ) + geom_line(aes(x = dates, y=min_press)) +

  scale_x_date(breaks = '10 years', labels = date_format("%Y") ,
    limits = as.Date(c('1849-01-01','2014-01-01'))) +
  labs(x='Year', y='Pressure (mb)',
    title='Minimum Pressure (mb)')

}
#
# make plot
#
presshist(tr)

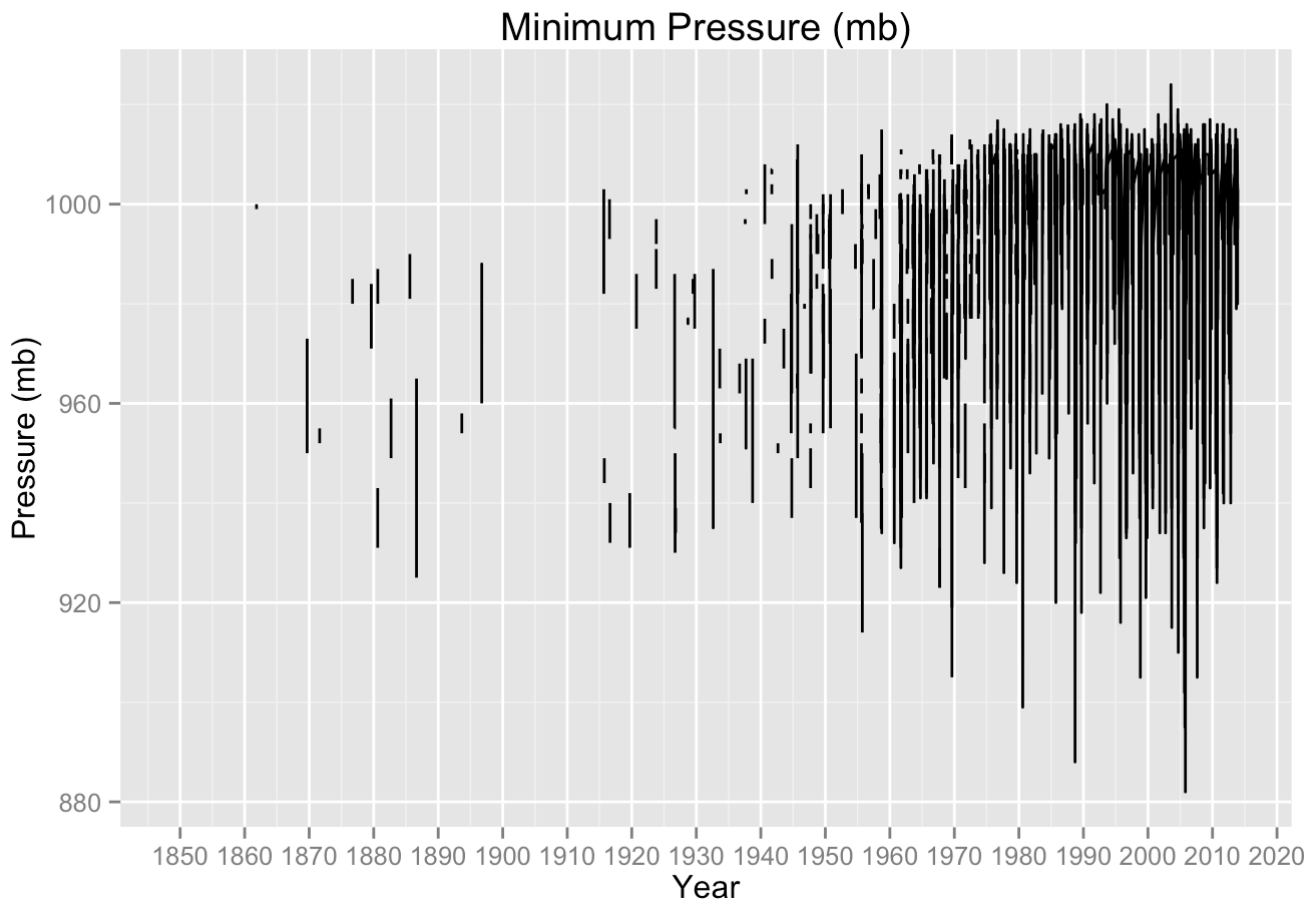
```

```

## Warning: Removed 127 rows containing missing values (geom_path).

```





In the early years of this data set before WW II roughly, there were few pressure measurements available. Measurements of pressure were generally made by ships of opportunity. So there is a lot of missing data. From 1960 and on, there were regular observations made by hurricane hunter aircraft; so the pressure data were more complete and reliable.

So looking at the trends from 1950 and on, I see 5 spikes of prominent low pressure, at 1955, 1969 (Hurricane Camille), 1981, 1997 (Andrew), and 2007 (Katrina – over the water). Now these comprise an extremely small sample, so it would be hard to argue for a trend, but it certainly raises the idea that a warmer climate might support additional record-breaking hurricanes in the Atlantic. So far I would say there is only a little evidence that the risk associated with major hurricanes is increasing as the world warms.

These results indicate a gap between public and media perception that hurricanes definitely are getting worse and more frequent as a result of climate change. For example, media sources often point to Hurricane Sandy and Katrina as an example of what will happen more often in the future. (Even the current Hollywood movie “Into the Storm” made a reference to Sandy and Katrina in this regard!) The results shown here indicate that hurricane activity has always been a significant risk to the United States, and will continue to do so, and that there is no “definite” association between climate change and hurricane activity.

Now the Atlantic basin is not the only place in the world with hurricanes! The Pacific and Indian Oceans, being much larger and warmer than the Atlantic, spawn many more intense hurricanes. So an idea for additional research is to repeat this analysis using track data from Pacific and Indian oceans.

## Reflection

I was interested in analyzing this particular set of data, as I have, in previous years, been a professor of meteorology, and taught classes on tropical meteorology. Though the HURTRAK data set was always available for use, interactive tools like R were not around in the 1980s and 90s when I was teaching and doing research on hurricanes. I would

have had to develop all the code myself in C, after buying a plotting package. And the computing power I had on the desktop then was nothing compared with the new Mac I have and used for this analysis.

So I undertook the effort to parse the HURTRACK data using R, partly as an experiment to see how well this language works. I found it was tougher than expected, because I was unfamiliar with the syntax, the function calls, and the overall behavior of R.

A lot of the coding went quickly – instead of writing computation loops, the automated vectorization made coding simpler. Instead of writing loops to check the values of variables, I discovered the `ifelse()` function, and used that extensively.

I was pleased with how easy it is to produce a simple map of the world in R. And plotting points and lines was easy and straight forward. But I spent a lot of time looking for advanced plotting packages that would make my heat map for me right away. After wasting a lot of time, I realized I just had to code it myself and use line and point functions.

But I got hung up multiple times when code running in scripts was accessing global variables I discovered that using the `<-` - assignment, instead of the `<-` limits the scope of a variable, and I started using that in functions on a regular basis to limit the namespace pollution.

On the other hand, it was nice to be able run code in a function, and access the results from the console.

I used the interactive debugger on several occasions, and discovered differences between the treatment of NA and NaN for missing data. NaN qualifies as a numeric value; but NA does not qualify as a numeric value. I got error messages regarding non-numeric values being present in a vector when I used NA to flag missing data, but those errors went away when I flagged missing data with NaN.

I ran into bugs converting between string and numeric formats, got return values in lists or arrays in different contexts, and found some functions were sensitive to the difference between a list and an array. So there is obviously is a lot of theory in R that I need to understand if I am going to use R going forward in my career.

At the end of it all, R programming is not a whole lot easier than Java or C, except for really simple data structures. My data structures were moderately complicated; so it still took a lot of time.

The results were very satisfying to me. I solved the problem of “the spaghetti plots” – which is what the tropical meteorologists call those track plots that go all over the place, and you really cannot analyze the plots very well. I showed how the heat map makes it easy to see where the action and risk of hurricanes really is. I was also pleased with the final plot, of the history of minimum pressure. If indeed this is a trend, it shows up in less than 1/1000 of all the observations. So this may be just a chance result. Detailed statistical modeling would be a good follow up research project.