Final Project Report

Hye Soo Choi

August 14, 2015

Data source

In this project, we analyzed the data about storms that hit North Atlantic(NA) and East Pacific(EP) from 1980 to 2010, in terms of location, wind speed, surface pressure, date of occurence, and duration. We used following storm data for these two basin :

- (ftp://eclipse.ncdc.noaa.gov/pub/ibtracs/v03r06/wmo/hurdat_format/basin/Basin.NA.ibtracs_ hurdat.v03r06.hdat),
- (ftp://eclipse.ncdc.noaa.gov/pub/ibtracs/v03r06/wmo/csv/basin/Basin.NA.ibtracs_wmo.v03r06. csv),
- (ftp://eclipse.ncdc.noaa.gov/pub/ibtracs/v03r06/wmo/csv/basin/Basin.EP.ibtracs_wmo.v03r06.csv),

which can be obtained at the **International Best Track Archive for Climate Stewardship (IBTrACS)** (ftp://eclipse.ncdc.noaa.gov/pub/ibtracs/v03r06/wmo/cxml/year).

Data cleaning

A cleaned data.frame, *storms*, was created by mainly using regular expressions to extract various pieces of data. After extracting the variables of interest and changing the variable formats, the final data include the following variables:

- id: storm id (e.g. 1)
- date: date (e.g. 11/13/2009)
- duration: duration in day (e.g. 3)
- name: name of the storm (ALLEN)

A cleaned data.frame, *tracks*, was created by mainly using regular expressions to extract various pieces of data. After extracting the variables of interest and changing the variable formats, the final data include the following variables:

- id: storm id (e.g. 1246)
- date: date (e.g. 07/18/1980)
- period: 00h, 06h, 12h, 18h (e.g. 00h)
- stage: type of stage (e.g. cyclone)
- lat: latitude (e.g. 30.5)
- long: longitude, negative for west (e.g. -76.5)
- wind: wind speed in knot (e.g. 25)
- press: surface pressure in mb (e.g. 1020)

The following displays the first 6 observations in the storms and tracks data.

```r
head(storms, 6)
```

```
##   id       date days    name
## 1  1 06/25/1851    4 MISSING
## 2  2 07/05/1851    1 MISSING
## 3  3 07/10/1851    1 MISSING
## 4  4 08/16/1851   12 MISSING
## 5  5 09/13/1851    4 MISSING
## 6  6 10/16/1851    4 MISSING
```

```r
head(tracks, 6)
```
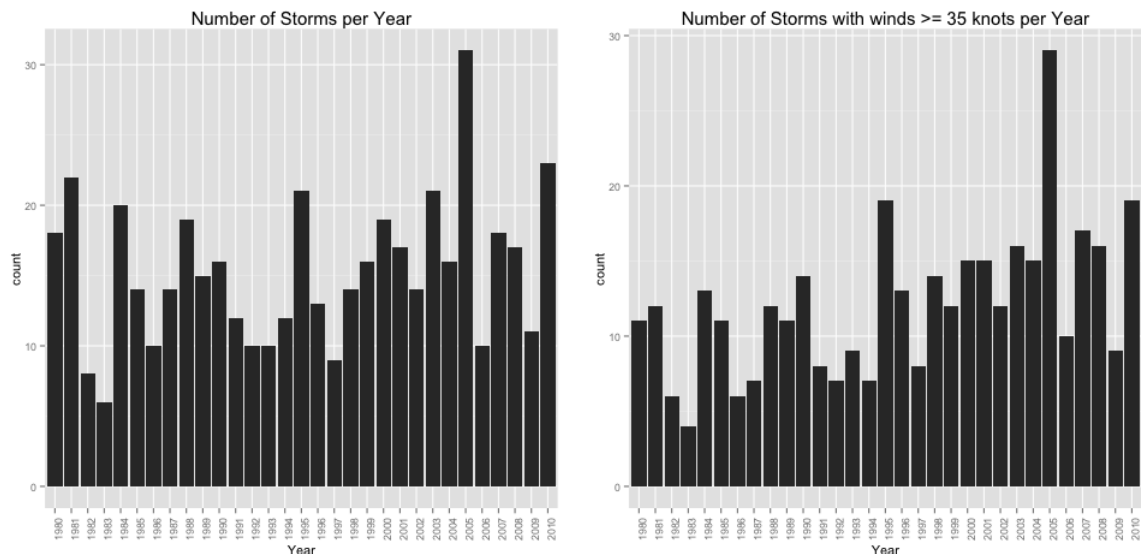
```
##   id       date period   stage  lat  long wind press
## 1  1 06/25/1851    00h cyclone 28.0 -94.8   80     0
## 2  1 06/25/1851    06h cyclone 28.0 -95.4   80     0
## 3  1 06/25/1851    12h cyclone 28.0 -96.0   80     0
## 4  1 06/25/1851    18h cyclone 28.1 -96.5   80     0
## 5  1 06/26/1851    00h cyclone 28.2 -97.0   70     0
## 6  1 06/26/1851    06h cyclone 28.3 -97.6   60     0
```
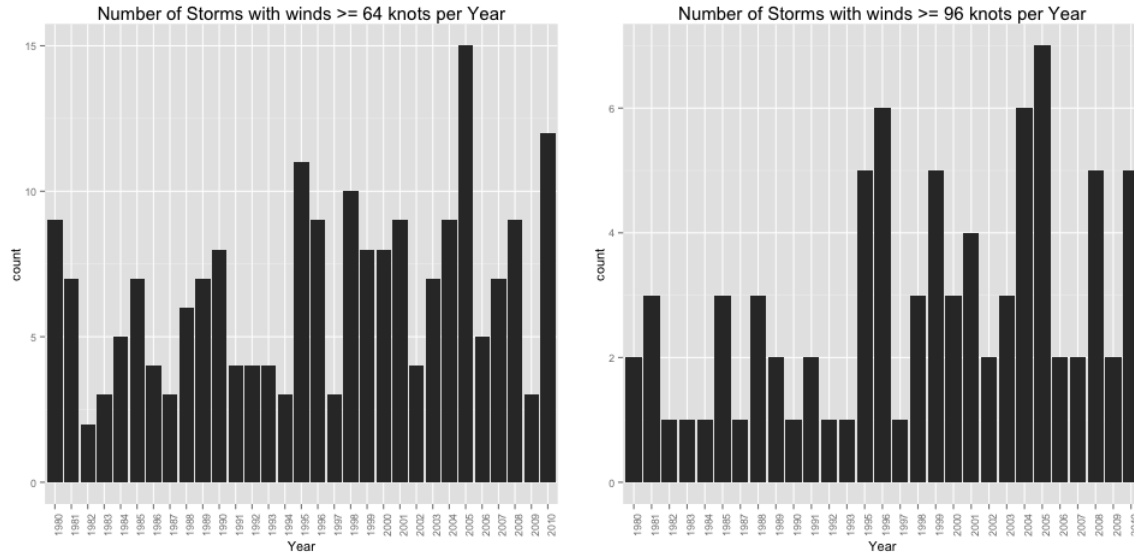

Data analysis

The first exploratory analysis was performed with respect to the number of storms per year. More specifically, we conducted exploratory anayasis on the following variables.

- Number of storms per year
- Number of storms per year with wind speed >= 35 knots
- Number of storms per year with wind speed >= 64 knots
- Number of storms per year with wind speed >= 96 knots

To understand the distribution, we produced barplot for each variable.

We can observe that storms were most frequent and violent in 2005.

We also examine the descriptive statistics of the variables. The following table presents the summary statistics of the four variables.

Table 1: Summary Statistic of annual number of storms by wind speed

	mean	sd	min	max	25%	50%	75%
	15.354839	5.263385	6	31	11.5	15	18.5
>= 35	12.161290	4.993975	4	29	8.5	12	15.0
>= 64	6.612903	3.116381	2	15	4.0	7	9.0
>= 96	2.862069	1.787339	1	7	1.0	2	4.0

The second exploratory analysis was performed with respect to the number of storms per month. More specifically, we conducted exploratory anayasis on the following variables.

- Number of storms per month
- Number of storms per month with wind speed >= 35 knots
- Number of storms per month with wind speed >= 64 knots
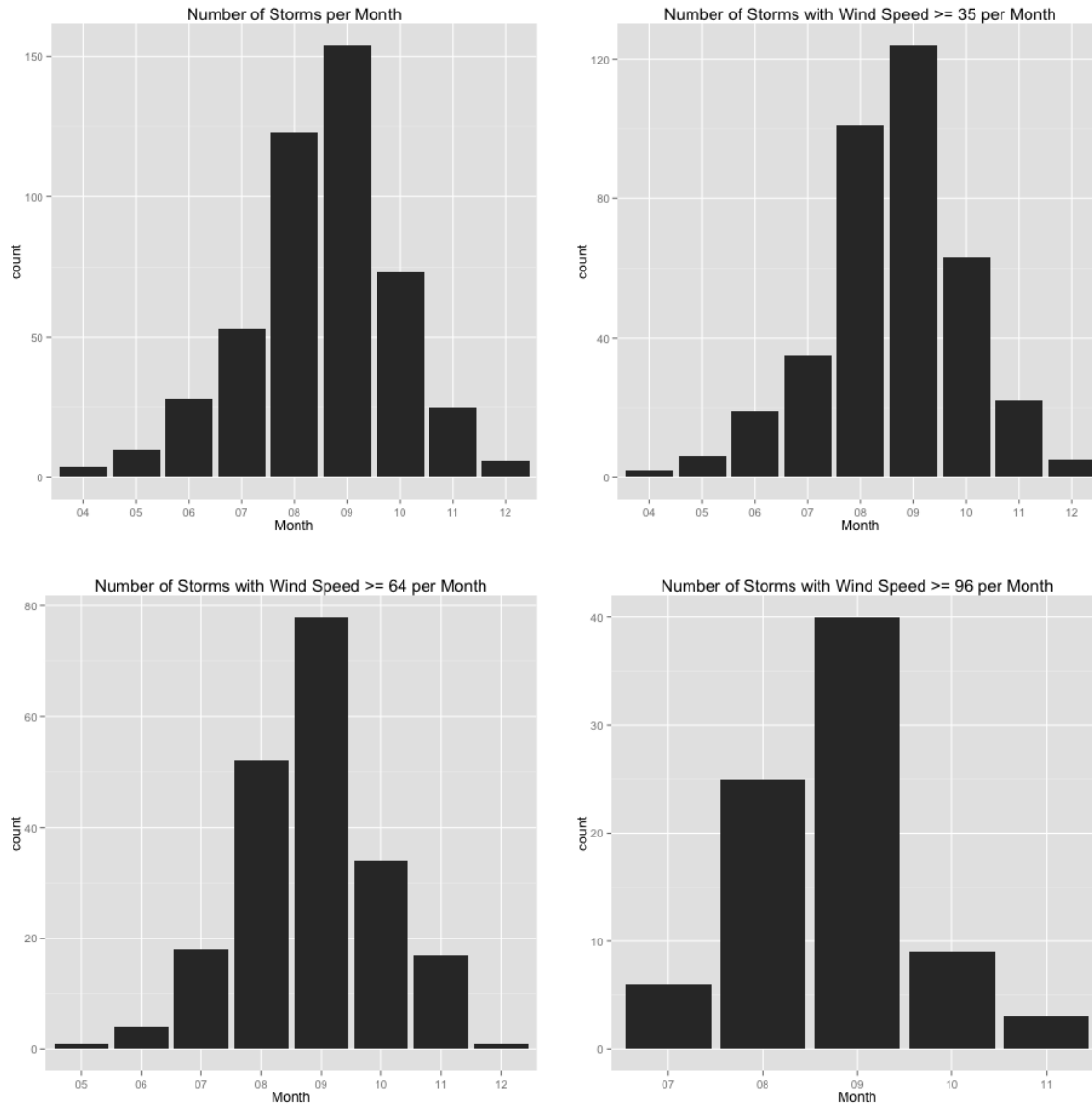- Number of storms per month with wind speed >= 96 knots

The follwing table presents the number of storms per month.

Table 2: Number of storms per month

month	count	count.35	count.64	count.96
4	4	2	NA	NA
5	10	6	1	NA
6	28	19	4	NA
7	53	35	18	6

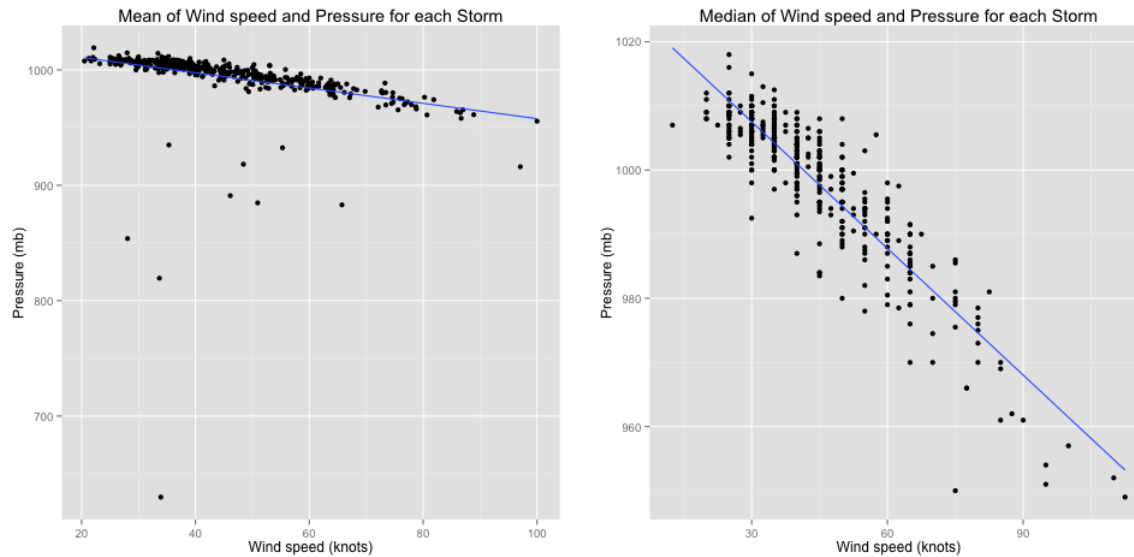| month	count	count.35	count.64	count.96
8	123	101	52	25
9	154	124	78	40
10	73	63	34	9
11	25	22	17	3
12	6	5	1	NA

To understand the distribution, we also produced barplot for each variable.

We can see that this graph is bell-shaped, and there has been most frequent and violent storms around september.

Regression Analysis

To estimate the relationship between wind speed and pressure for storms, we used linear regression. For each storm, mean(median) of wind speed and pressure is calculated and plotted, and then the linear regression line is drawn on the scatter plot of wind speed and pressure. To improve prediction accuracy, we removed storms whose mean(median) pressure is zero in our regression analysis.
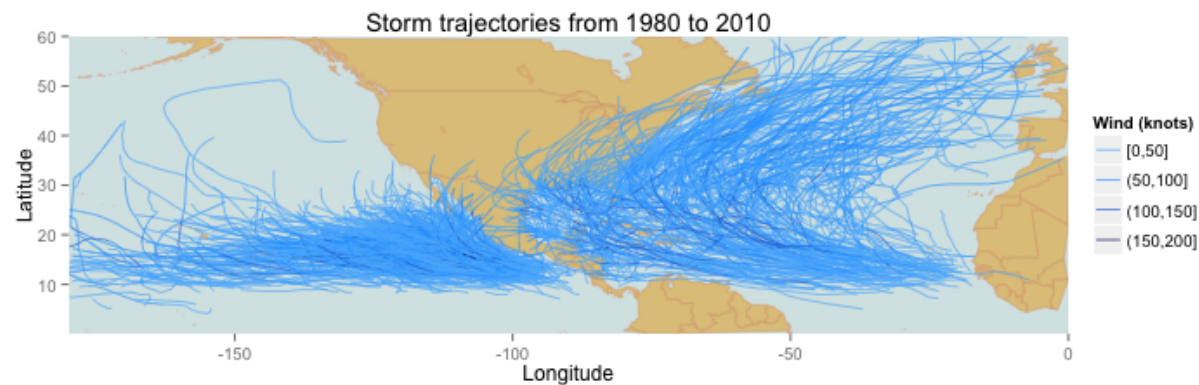

Although more outliers are present, residual values seem to be smaller in the scattor plot of mean wind speed and mean pressure than in the scatter plot of median wind speed and median pressure. We can observe that wind speed and pressure has strong negative linear relationship in both scatter plots.
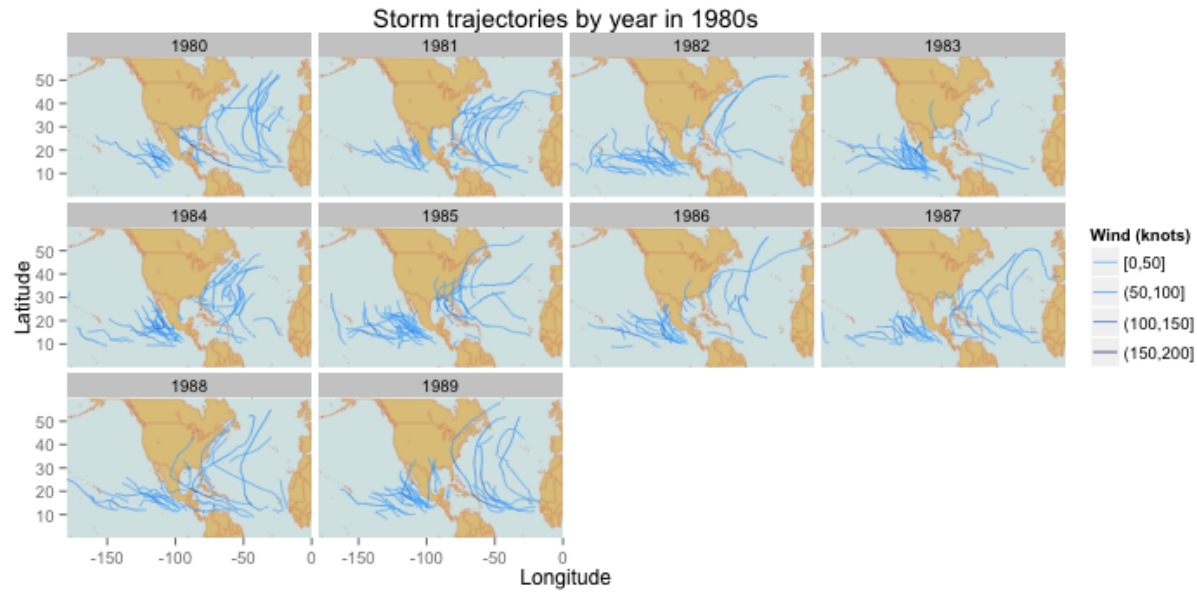

Visualization

The third part of the project consists of visualizing the trajectory of the storms in both the East Pacific (EP) and the North Atlantic (NA) basins, for the period 1980-2010.

We produced a world map with spatial distribution of storm activity. The color of data points reflects storm intensity; the higher wind speed was, the darker blue a point is colored on the map.
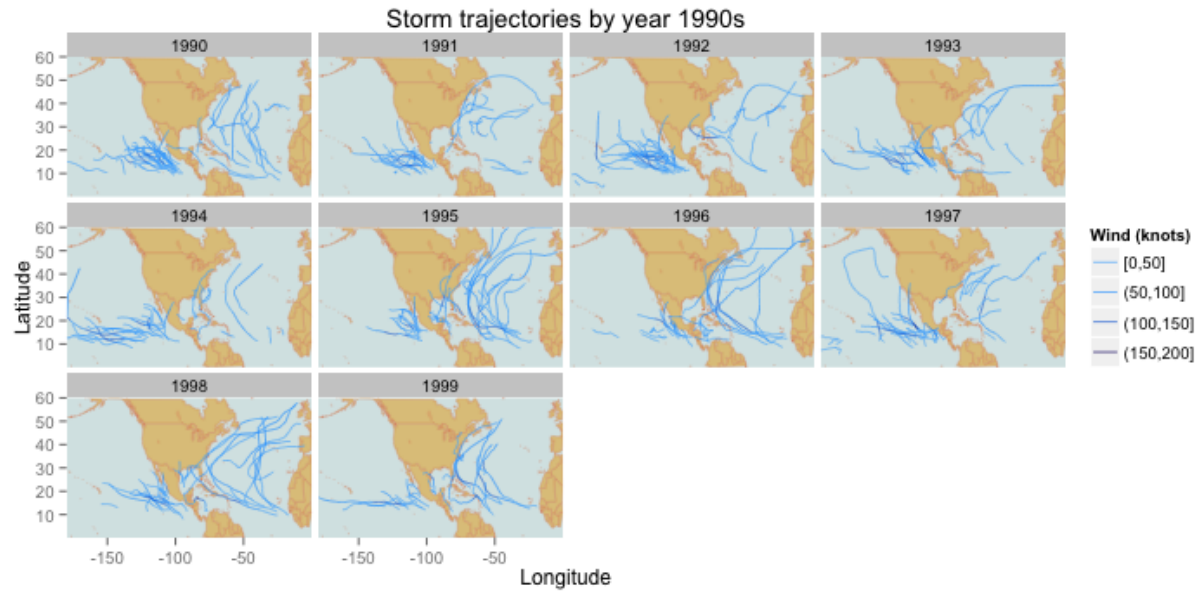
To be more specific, we draw world map with spatial distribution of storms in different decades: 1980s, 1990s, 2000s.
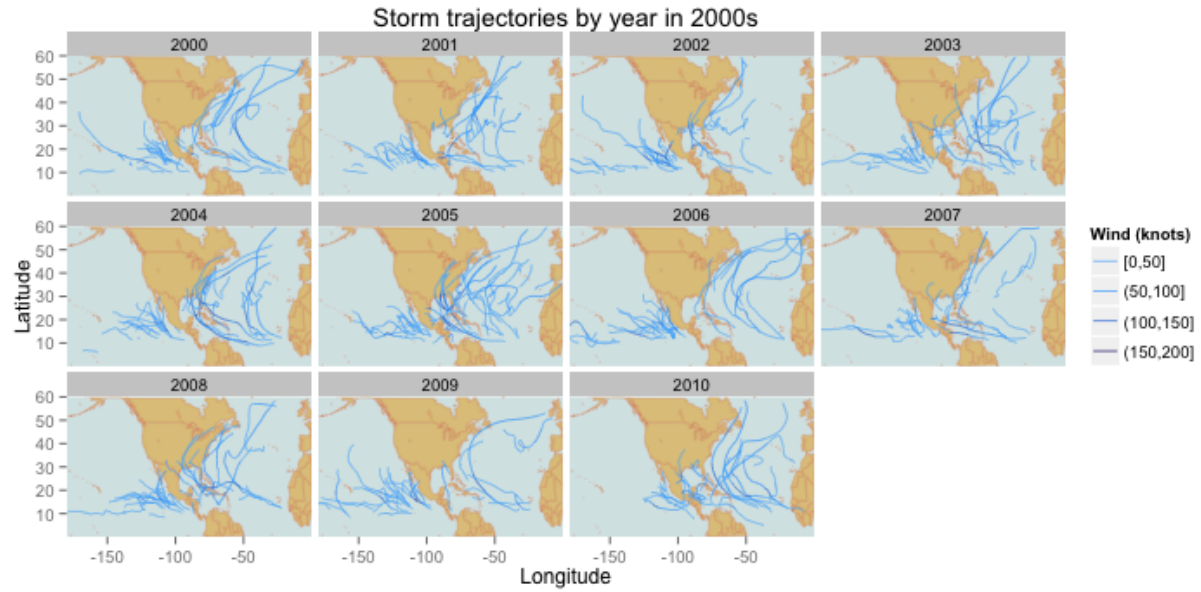
In 1980s, storm trajectories for each year is drawn as follows.
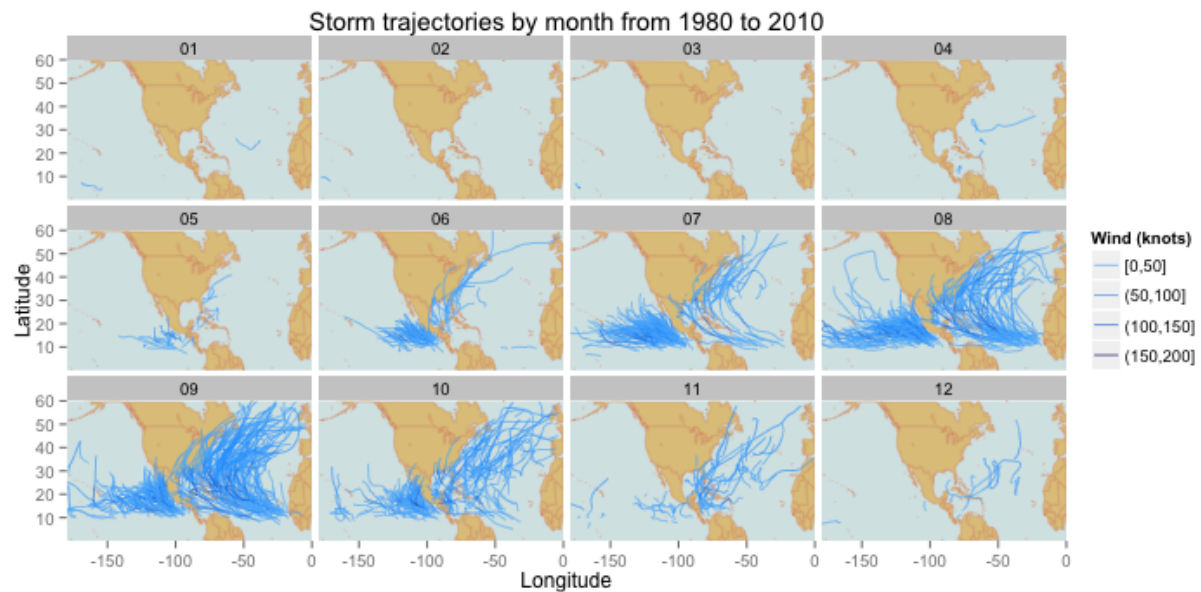

Storm trajectories by year in 1980s

In 1990s, storm trajectories for each year is drawn as follows.

Storm trajectories by year 1990s

In 2000s, storm trajectories for each year is drawn as follows.

Storm trajectories by year in 2000s

Although, the trajectories varies from year to year, We can see simlilar path of storms throughout decades:

- Across North Atalantic basin, storms generally move westward with a slight tend toward the north.
- Across East Pacific, storms generally move eastward with a slight tend toward the north.
- High wind speed, higher than 100 knots, usually occurs in latitude between 10° and 30°.

We also produced the storm trajectories for each month.

Storm trajectories by month from 1980 to 2010

As shown from the barplot of number of storms per month, there has been most frequent and violent in September, and secondly in August. In January, Febraury, and March, storms rarely occurs.