Homework7 Report

Hye Soo Choi

August 2, 2015

Data source

In this study, we analyzed the data about storms that hit the world in 2010, in terms of location, speed, pressure, date of occurence, and duration. We used the storm data of 2010 (Year.2010.ibtracs_wmo.v03r06.cxml), 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 *XPATH* expressions from **XML** package, to extract various pieces of data. After extracting the variables of interest and changing the variable formats, the final data include the following variables:

- name: name of the storm (e.g. ANJA)
- date: date (e.g. 2009-11-13)
- time: time (e.g. 06:00:00)
- latitude: latitude (e.g. -9.50)
- lat_deg: latitude degrees (e.g. "N")
- longitude: longitude (e.g. 72.50)
- lon_deg": longitude degrees (e.g. "E")
- press: pressure (e.g. 1006.0)
- wind: wind speed (e.g. 0.0)

The following displays the first 6 observations in the data.

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

```
##   name       date     time latitude lat_deg longitude lon_deg press wind
## 1 ANJA 2009-11-13 06:00:00     -9.5       N      72.5       E  1006    0
## 2 ANJA 2009-11-13 12:00:00    -10.2       N      71.9       E  1004    0
## 3 ANJA 2009-11-13 18:00:00    -11.1       N      71.4       E  1002   25
## 4 ANJA 2009-11-14 00:00:00    -11.9       N      71.1       E   999   28
## 5 ANJA 2009-11-14 06:00:00    -12.5       N      70.9       E   996   33
## 6 ANJA 2009-11-14 12:00:00    -12.8       N      70.7       E   992   40
```

Data analysis

An exploratory analysis was performed with respect to the pressure, wind speed, and duration of storms. We first examine the descriptive statistics of the variables. The following table presents the summary statistics of the three variables.

	mean	sd	min	max	25%	50%	75%
pressure	987.94	60.12	0.00	1016.00	985.00	998.00	1004.00
speed	42.33	25.88	0.00	140.00	25.00	35.00	55.00
duration	7.31	4.19	2.00	24.00	4.00	7.00	9.00

Table 1: summary statistics of pressure, speed, duration

To understand the distribution, we also produced histogram and boxplot for each variable.

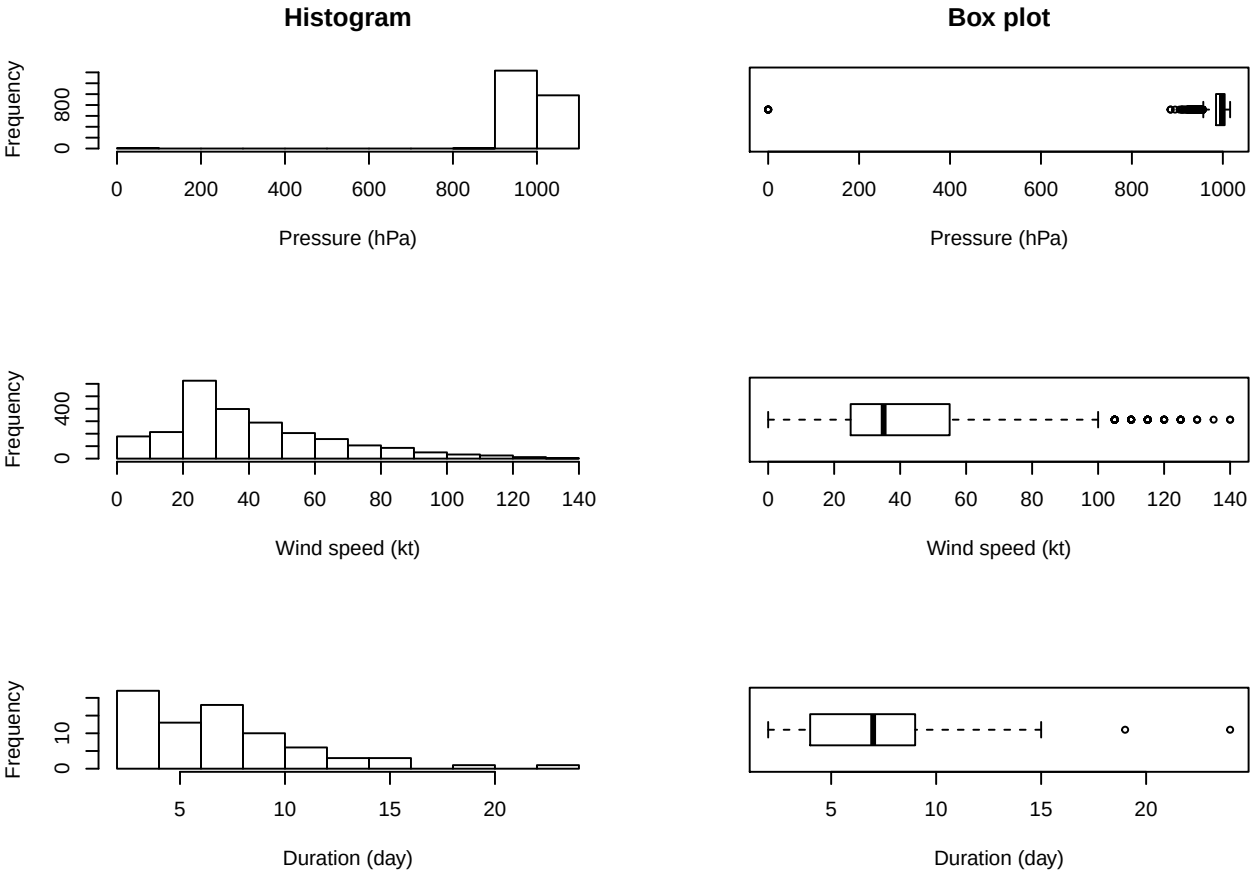

Figure 1: Histogram and Box plot of pressure, wind speed, duration

We can observe that the pressure is skewed to left, and the wind speed and duration are skewed to left.

In addition, we also found several informative factors of the storms:

- Total number of storms in 2010: 71

- Number of storms with winds >= 35 knots (tropical storms): 74
- Number of storms with winds >= 64 knots (hurricanes): 40
- Number of storms with winds >= 96 knots (major hurricanes) : 16
- Number of storms on the north hemisphere: 52
- Number of storms on the south hemisphere: 25
- Frequency table of data points per month (month in words) in the world, northern hemisphere, and southern hemisphere:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
World	142	173	212	80	96	201	108	305	519	240	125	225
Northern	122	60	171	80	87	50	53	98	181	111	84	202
Southern	20	113	41	0	9	151	55	207	338	129	41	23

Table 2: Frequency table of data points per month in the world, and nothern and southern hemisphere

- Name of storm that lasted most days : TWO
- Name of storm with maximum wind speed and the maximum speed value: BLAS and 140
- Name of storm with minimum pressure and the maximum pressure value: HUBERT, PHET and 0

Data visualization

For the visualization of the data, we produced a world map with spatial distribution of storm activity in 2010. The color of data points reflects the intensity of each storm activity; the higher wind speed was, the darker blue a point is colored on the map.

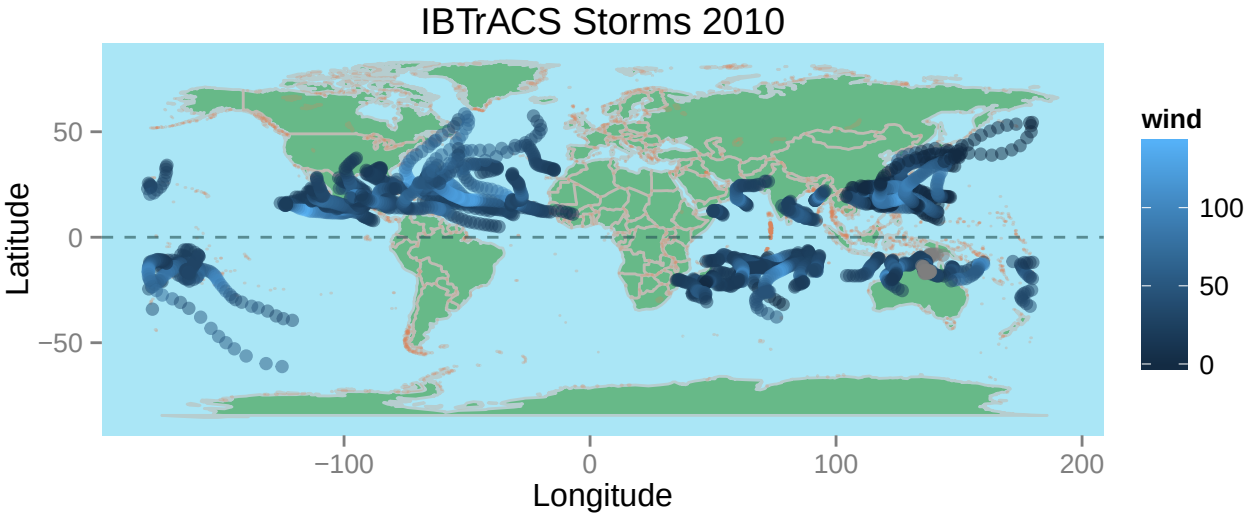

Figure 2: Spatial distribution of storm activity in 2010