

# Wind Analysis Of Northern Yucatan Coast for Kiteboarding

*Roque Arcudia, Arshia Razavi, Eda Uysal*

*August 21, 2017*

## Motivation

Kiteboarding is an adventure water sport that consists of using a large steerable kite to pull yourself around on a board on top of the water. The kites are controlled through the bar and lines that connect to your harness, which is worn around your midsection. The kite can be flown specific patterns to generate power, and can even be used to jump high in the air. And when we say jump, we really mean it. Professional kiteboarders can jump 50 or 60 feet high and can easily float for hundreds of feet when the conditions are right. That's nearly 10 seconds of air time!

Kite boarding is the fastest growing water sport in the world. The key to this sport is the Wind. This is what motivated us to do our analysis on Wind. Wind being the key element of kiteboarding, there are a lot of details that come into play for a safe and good ride. Wind direction and wind power are the two most important elements for kiteboarders. The wind direction to be on-shore or side on-shore and the power to create enough power and acceleration for kiting. Wind power is normally measured in mph or Knots. A good kite boarding day starts at 15 mph. That being said, too much power is not safe. Meaning that smaller size kites are used for more powerful winds. Normally kiteboarders look at increments of wind to choose their kites. For a 15 - 20 mph, a 12 meter kite does a great job. For stronger winds between 20 - 25 perhaps a 9 meter is good for an average weighted person. for over 25+ mph depending on the skill level smaller kites should be used. Below is a good reference to learn more about winds and kiting: <http://www.gokite.com.au/resources/education/understanding-the-wind/>

The purpose of this analysis is to find the best windy days that you can travel to Northern Yucutan for your kiting vacation.

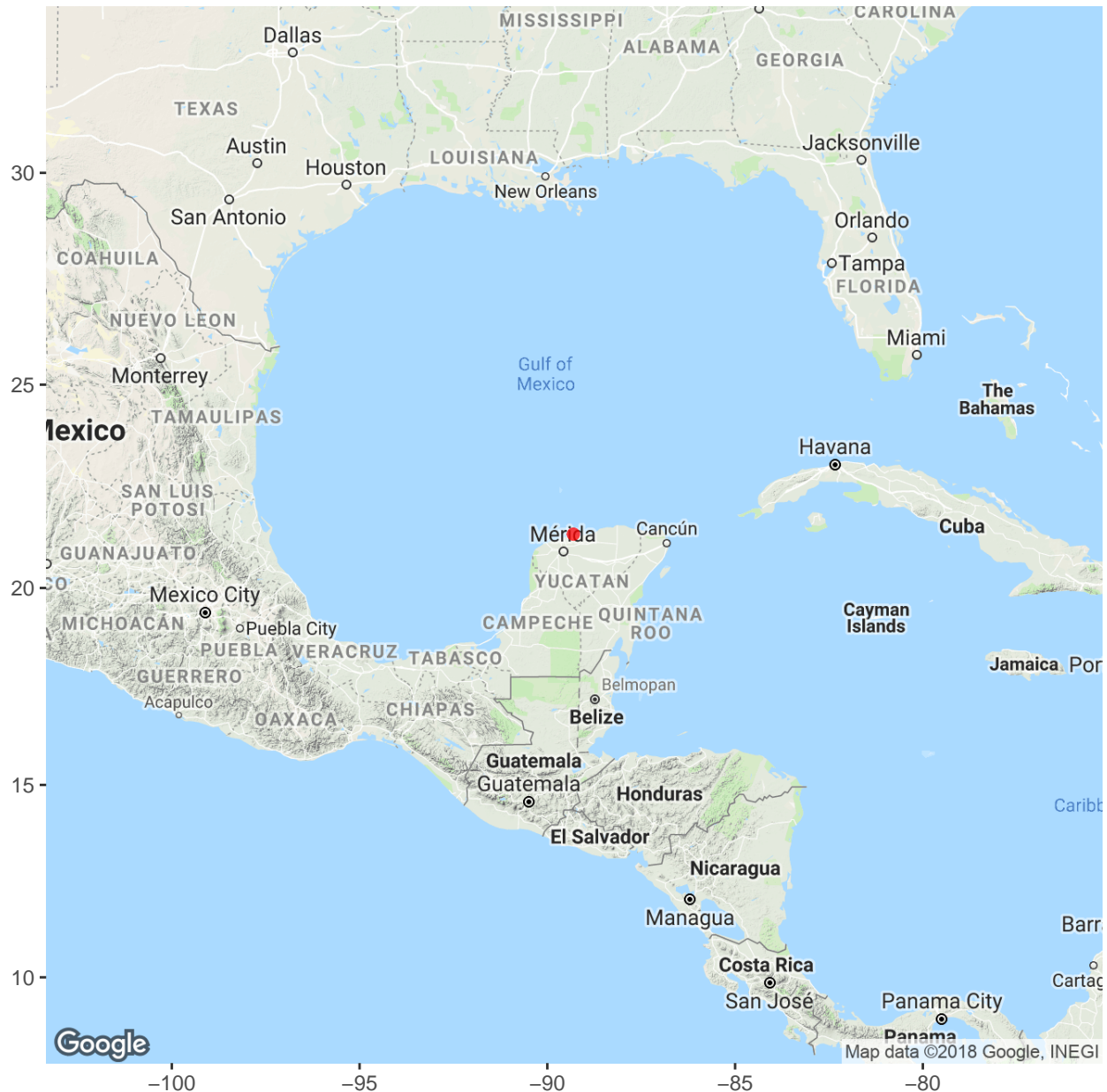
Happy Kiting!

## Data

### Data Sources

Original source is the CINVESTAV weather station (Mexican research center), ID IYUCATNT2, located in Telchac Yucatan. Coordinates are latitude 21.341108 and longitude -89.305756.

## Location of Station IYUCATNT2



We gathered the data from the weather underground website (<https://www.wunderground.com>). For instance the following URL will provide a close to CSV format (Except for some html tags inserted) for the day of January 1st 2016:

<https://www.wunderground.com/weatherstation/WXDailyHistory.asp?ID=IYUCATNT2&month=1&day=1&year=2016&format=1>

The project can be run in online mode where it will pull the data from weather underground website (<https://www.wunderground.com>) or it can run in offline mode where the data is pulled from a local cache saved in csv. Offline mode is strongly recommended since the online access relies on weatherData library that we had to patch for a bug plus weather underground data becomes unavailable from time to time. Instructions will be provided in the appendix.

## Summary of the Data

We analyzed a total of 5 years of data from 2012 to 2016. The information was obtained from weather underground website giving a single web page summary per day. This is an extract from April 18, 2016:

Time	WindDirection	WindDirectionDegrees	WindSpeedMPH	WindSpeedGustMPH
2016-04-18 00:00:00	East	101	7	7
2016-04-18 00:05:00	ENE	70	7	12
2016-04-18 00:15:00	East	82	10	11
2016-04-18 00:20:00	East	87	9	14
2016-04-18 00:30:00	East	100	6	6
2016-04-18 00:35:00	SSE	147	7	7

There are many more columns related to pressure, temperature, dew point, etc. but for this analysis we only focused on the following:

*Time* : The data we are dealing with is a time series. This is the timestamp at which the sample is taken.

*WindDirection* : Cardinal direction where the wind is coming from. Not really used for analysis but just for display.

*WindDirectionDegrees* : This is the wind direction in degrees. The cardinal directions map as follow: North is 0°, East is 90°, South is 180° and West is 270°.

*WindSpeedMPH* : Wind speed measured in miles per hour (mph).

*WindSpeedGustMPH* : Wind gust measured in miles per hour (mph).

During the period of 5 years a total of 1740 records were read from the web and 87 were missing from the server. This gives a total of 329115 measurements to analyze.

This is the summary of the data from the raw data set:

Time	WindDirectionDegrees	WindSpeedMPH	WindSpeedGustMPH
Min. :2012-01-01 00:04:00	Min. :-737280	Min. :-999.900	Min. :-999.00
1st Qu.:2013-05-28 00:32:30	1st Qu.: 51	1st Qu.: 5.000	1st Qu.: 4.00
Median :2014-08-22 06:20:00	Median : 91	Median : 9.000	Median : 10.00
Mean :2014-08-09 17:35:51	Mean : -1936	Mean : 9.342	Mean : -90.73
3rd Qu.:2015-11-10 01:11:00	3rd Qu.: 128	3rd Qu.: 15.000	3rd Qu.: 15.00
Max. :2016-12-31 23:52:00	Max. : 32767	Max. : 41.000	Max. : 46.00

## Measurement Cleanup

We could easily spot there are negative values for the 3 columns of interest. Those are impossible values for those columns, further analysis showed that's the way the server reports the NA values. In case of wind direction there is also very big positive values greater than 360. When the wind speed is zero the server sometimes reports the direction as a negative value but when the direction reading is missing it reports it as a very high positive. The maximum numbers for wind speed and gust seem to be on a reasonable range so we consider them good readings.

Finally there are some entries in the data set that have duplicated Time values. They are not exactly duplicated readings since they could have other values in the remaining columns. It looks like sometimes some readings that were intended to happen minutes apart were recorded in the same timestamp. Instead of trying to guess the correct timestamp we simply deleted them, there were only 57 entries in this situation

plus future cleanup will minimize the effects of deleting them.

This is a summary of the data cleanup.

Condition	Action
WindDirectionDegrees < 0	Convert to NA
WindDirectionDegrees > 360	Convert to NA
WindSpeedMPH < 0	Convert to NA
WindSpeedGustMPH < 0	Convert to NA
duplicated(Time)	Delete

Although converting to NA doesn't sound like a good cleanup strategy it will make sense in the next step.

## Temporal Data Cleanup

The data is a time series which should be indexed by the Time column. The data set doesn't have any information of the time zone which becomes important for future analysis. We had to force such information to the Time column setting the timezone to "America/Mexico\_city"

The minimum difference between readings of the same day is 5 minutes but there are situations when the reading for the next 5 min is not taken or the period shifts. The following table shows the 5 most common intervals in the data set (in seconds):

Intervals	Freq
300	164962
600	143652
540	7490
360	6481
900	2934

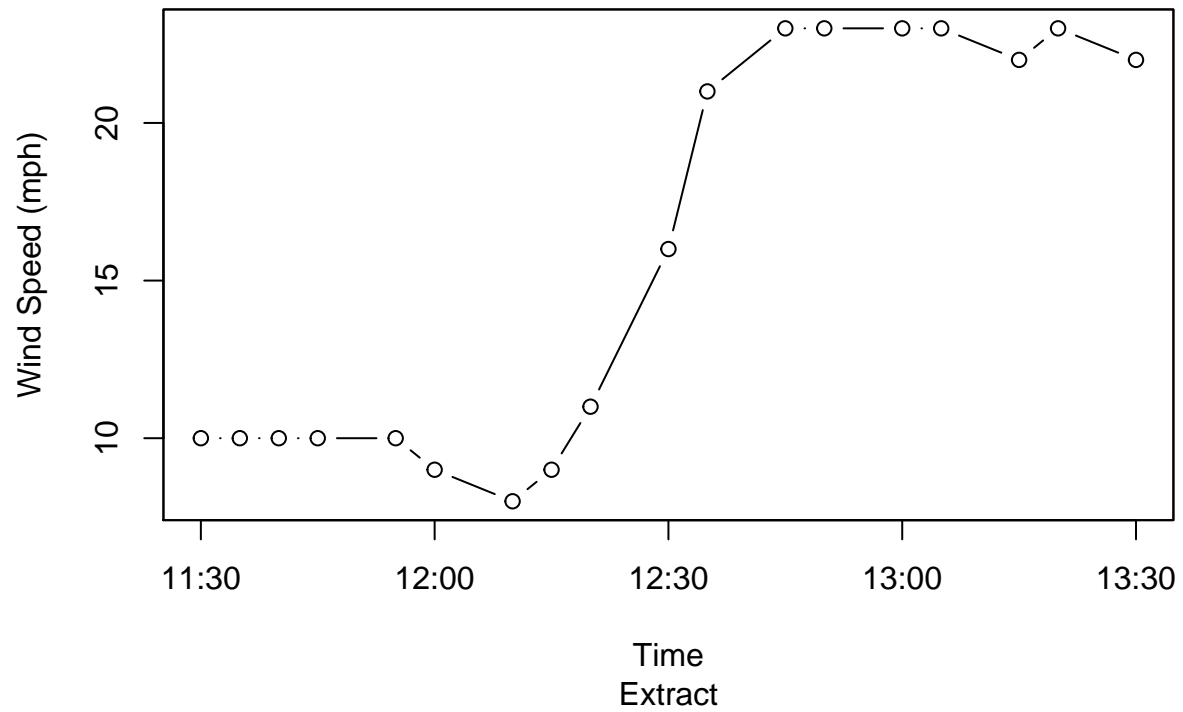
In order to make the data set more manageable for future processing we decided to make each day evenly sampled at 5 minutes interval adding interpolated values in the case of missing timestamp.

Since the data set is a temporal series we converted the data to zoo objects in R and executed the following steps for interpolation for each day:

1. Create a new empty data set (zoo) with only a time series starting with the first timestamp of the day we want to interpolate, ending with the same timestamp as the original data set and equally spaced by 5 minutes.
2. Merge the new empty data set with the old original one for the date as a full outer join creating NA readings for the timestamps that didn't exist.
3. Use the operation na.approx on the merged zoo object which will result in approximating the NA values using interpolation between neighboring readings.
4. Merge again the empty dataset with the merged data set as a natural join. This will eliminate any irregular reading taken outside the 5 min sampling rate.

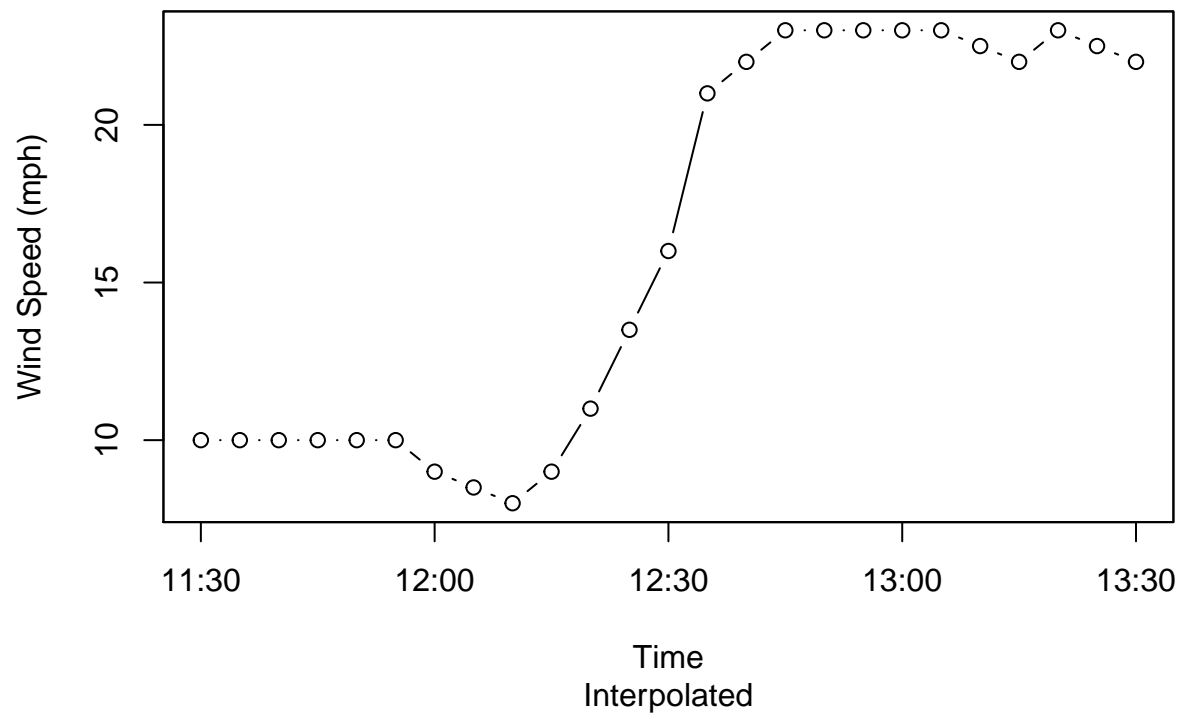
Let's take a look at an extract of samples around noon from April 18, 2016:

### Wind Speed for April 18, 2016



The next graph is the same period of time with the interpolated samples added:

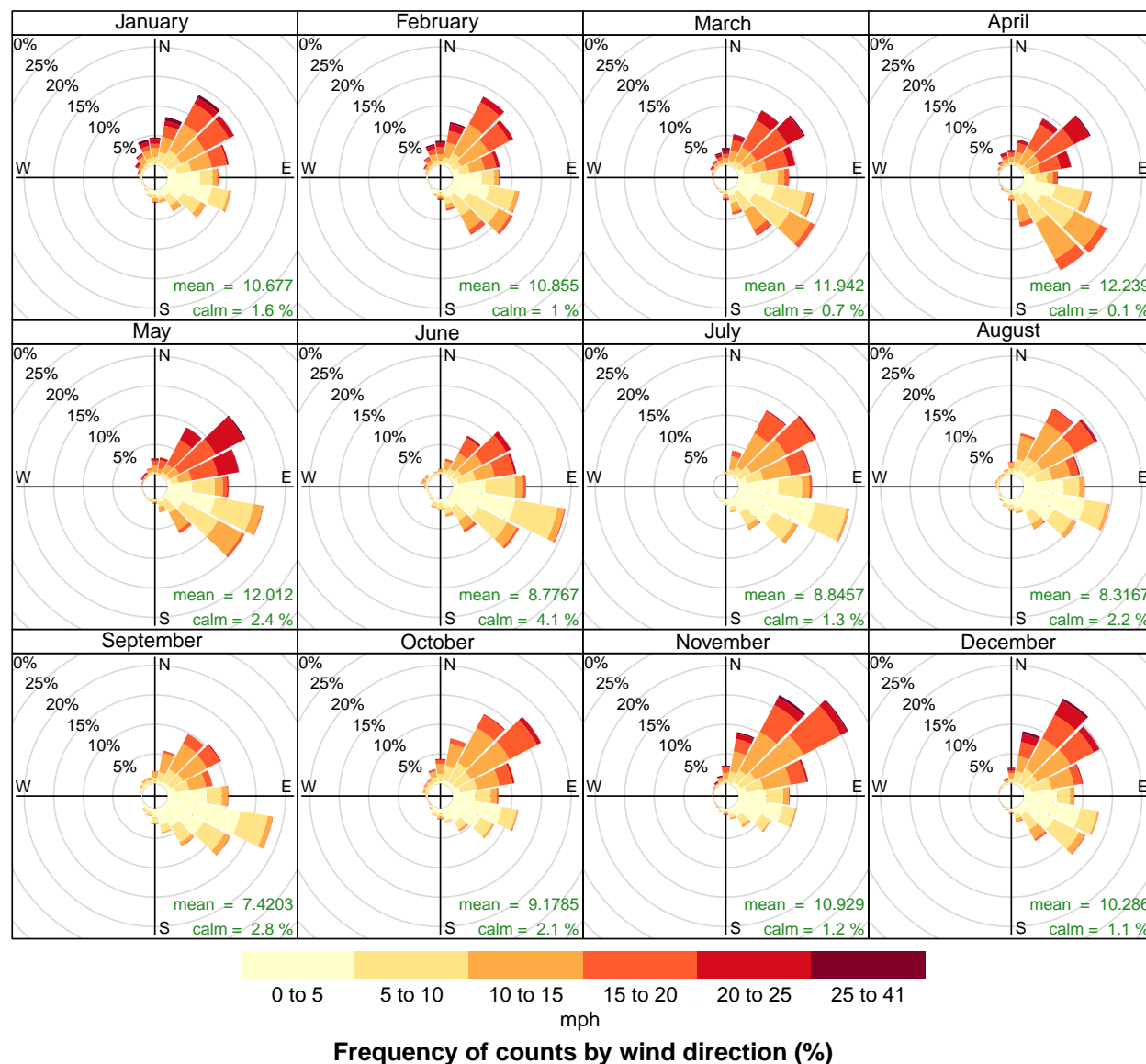
### Wind Speed for April 18, 2016



# Analysis

## Whole Data Set Analysis

The first analysis done after cleaning the date was to get a windrose diagram to see the wind speed and direction distribution for the 5 years we are analyzing. A windrose diagram is just a histogram in polar coordinates. It is divided by month since we are interested on finding the good season for kiting.



Visually we can observe that the months of August and October look like the weakest. The winter-spring months seem to have the highest winds overall. May seems to be a good solid month.

There are 2 dominant wind directions north-east and south-east. NE winds are stronger overall compared to the SE ones. This is good news since the NE direction creates a side-onshore direction excellent for kiting in the north shore.

## Daily Summary

The information from the wind rose is a good quick outlook of how the winds distribute along the year but it doesn't answer the question of how many good days for kiting a month has.

Instead we went for a more “visual” approach. Everyday is scanned to look for a period of at least 2 hours of sustained winds in the ranges of 15, 20 and 25 mph. The fact that each day was resampled at 5 min intervals greatly simplifies this analysis since it becomes a counting problem rather than calculating time differences.

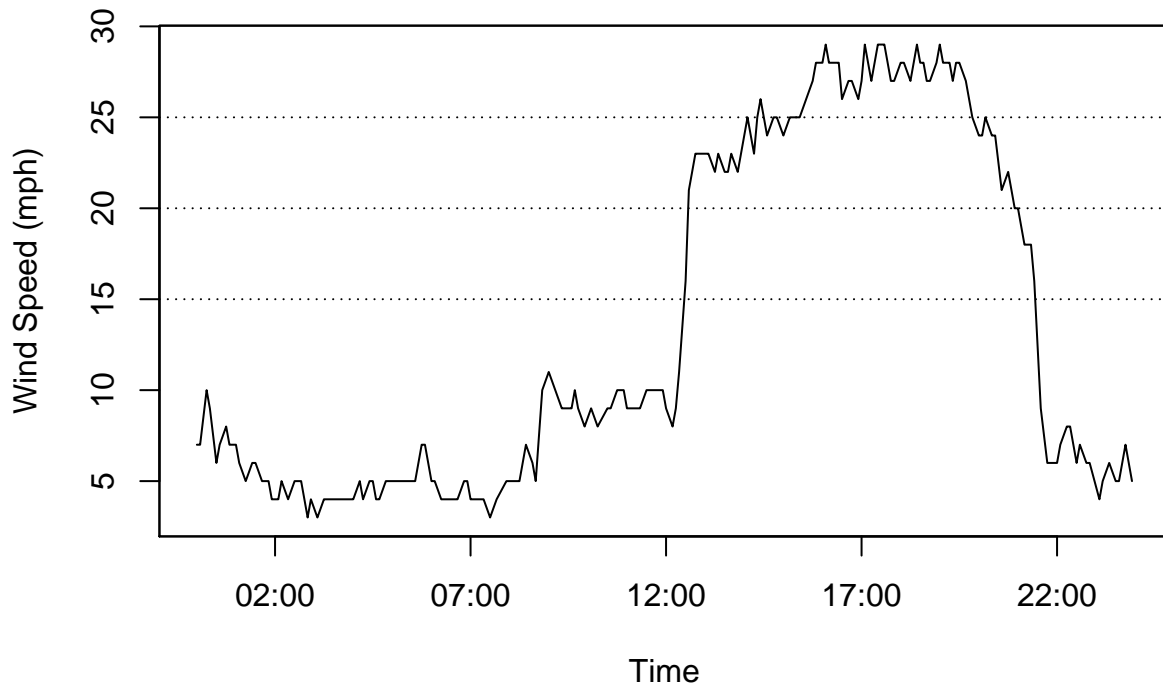
Kiteboarding is only practice during daytime, we removed the measurements that were taken (or interpolated) at times when there was no sunlight before computing the daily summaries. The sunset and sunrise times are calculated for each day using the `sunriseset` function from `maptools`.

This is a list of the computations performed using the column name in the final daily summary data frame:

*avgDlWindSpeedMPH*, *avgDlWindDirectionDegrees* : This is the average wind speed and direction during the daylight. The averages are not relevant for the final report but they provided a way to do early estimation.

*periodLength15*, *periodLength20*, *periodLength25* : These are the most important measurements. We look for the amount of time the wind speed was equal or greater than 15, 20 and 25 mph respectively. A time range is only good if we have at least 2 hours of contiguous wind over the desired speed otherwise the time is not counted (becomes zero). If there are multiple periods going over the 2 hours threshold then they are added together. Lets use again the example of April 18, 2016:

### Wind Speed for April 18, 2016



This one contains period of times where the wind speed is sustained for more than 2 hours being equal or greater than the three thresholds. Non zero values will be captured in the three of them.

*pseudoWindSpeed* : We capture here the highest sustained speed above the predefined thresholds. In the previous graph for April 18, 2016 this field will have 25 since there is a period of time that the wind blows for more than 2 hours at 25 mph or more.

*avgDlWindSpeedMPHGt15*, *avgDlWindDirectionDegreesGt15* : These are the average wind speed and direction of all the measurements greater than 15 mph. Only calculated when there was at least a period of 2 hours

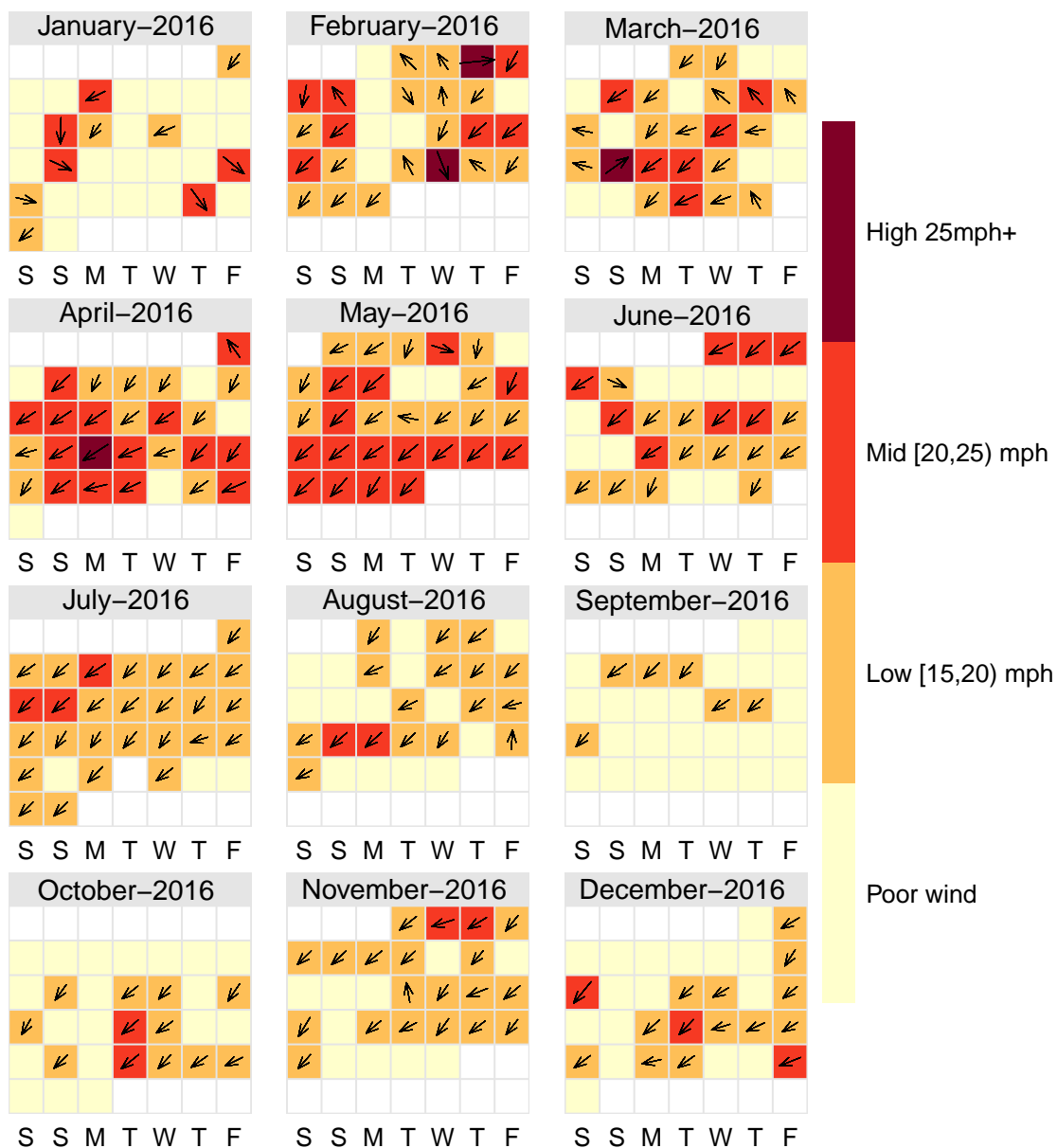
when the wind blows above 15 mph ( $\text{periodLength15} > 0$ ) otherwise they are NA. This measurements are useful for the arrow in the calendar plot.

### Calendar Plot for windy days

With this information we created a calendar plot where each day is classified into a no wind day, low wind, middle wind or high wind day based on the ranges of 15mph, 20mpg or 25mph and beyond. This is based again on the criteria of sustained wind for at least 2 hours.

Whenever there is a windy day ( $>15\text{mph}$ ) then we also get an arrow indicating the average wind direction of the windy range, this information comes from columns *avgDlWindSpeedMPHGt15* (size of arrow) and *avgDlWindDirectionDegreesGt15* (direction). For this report only the last year is printed, refer to the appendix for the remaining years.

### Kiting days in 2016

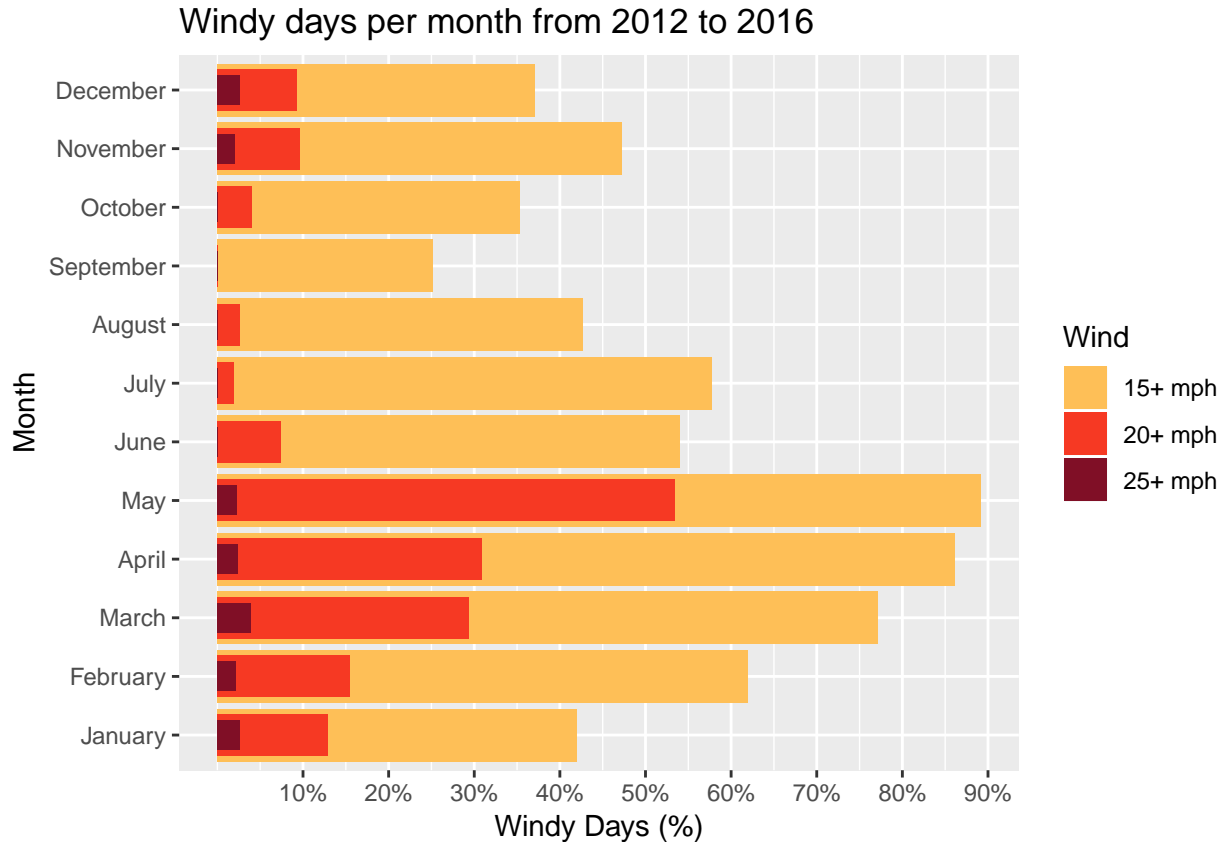




## Monthly Report

Finally we process the daily information to find out how many windy days we have had in the historic data per month. This is measured as a percentage.

The following overlapping bar charts were created as the final report:



## Conclusions

In the period from 2012 to 2016 the best month is May with 89% of windy days and 53% of windy days with at least 20mph speeds. This is followed by April with 86% of windy days and March with 77% of windy days.

The highest winds could typically be found in March with 4% of windy days with speeds of 25mph and above.

It is a safe bet to take a medium (9m) and a large (12m) kites since most of the windy days are in the range of 15mph to less than 25mph.

## Appendix

### A. How to Run

This is the list of source files and directories needed to run, all of them need to be present **in the same directory**:

**WeatherTest.R** : This is where most of the processing and plotting functions are implemented. It is included in the markdown documents.

**Report.Rmd** : This report.

**Final\_preso\_wind.Rmd** : Presentation.

**weatherc** : A directory containing all the cvs files with the daily information. It is organized in several subdirectories first the station ID, then the year and then the month.

To do the processing open the source file of this report and click on knit.

The library weatherData has been commented out from the code. This will make the code fail when the weather cache (directory weatherc) doesn't exist:

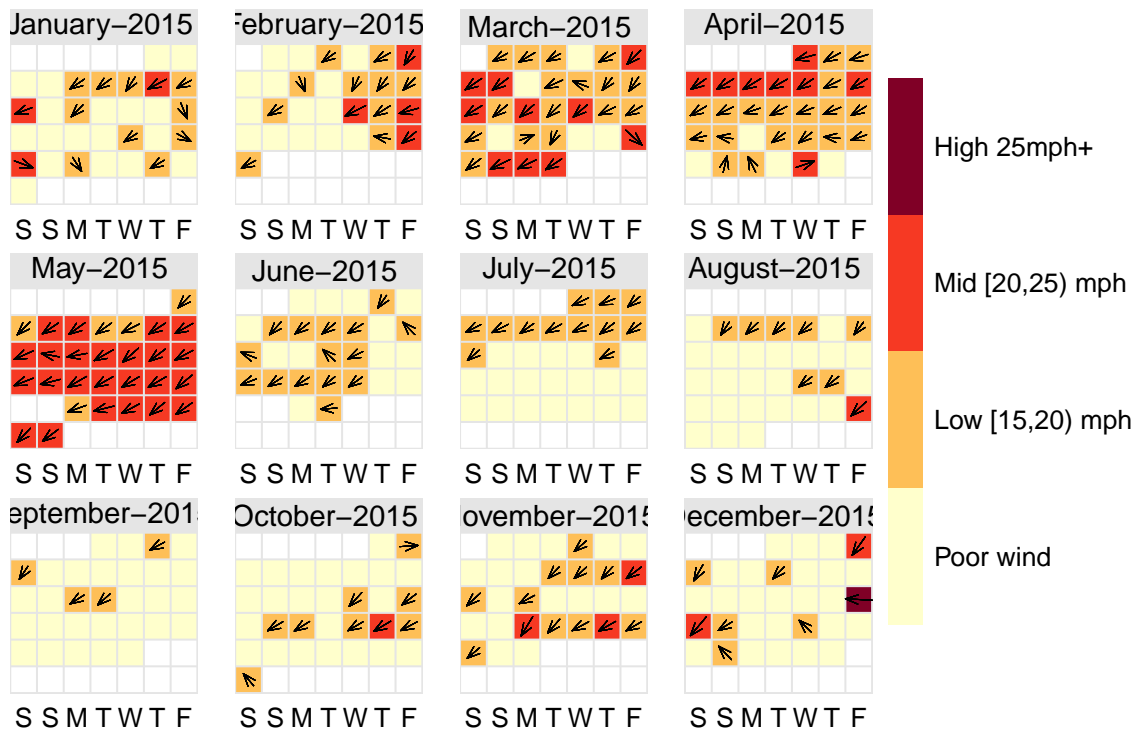
Quitting from lines 67-82 (Report.Rmd) Error in getPWSData(stationID, as.Date(sampleDate, "%B %d, %Y")) : could not find function "getDetailedWeather"

The current version from CRAN is old and it doesn't read the weather data from the page correctly. The one from github is better but it has an issue related to the hidden tags present in the page.

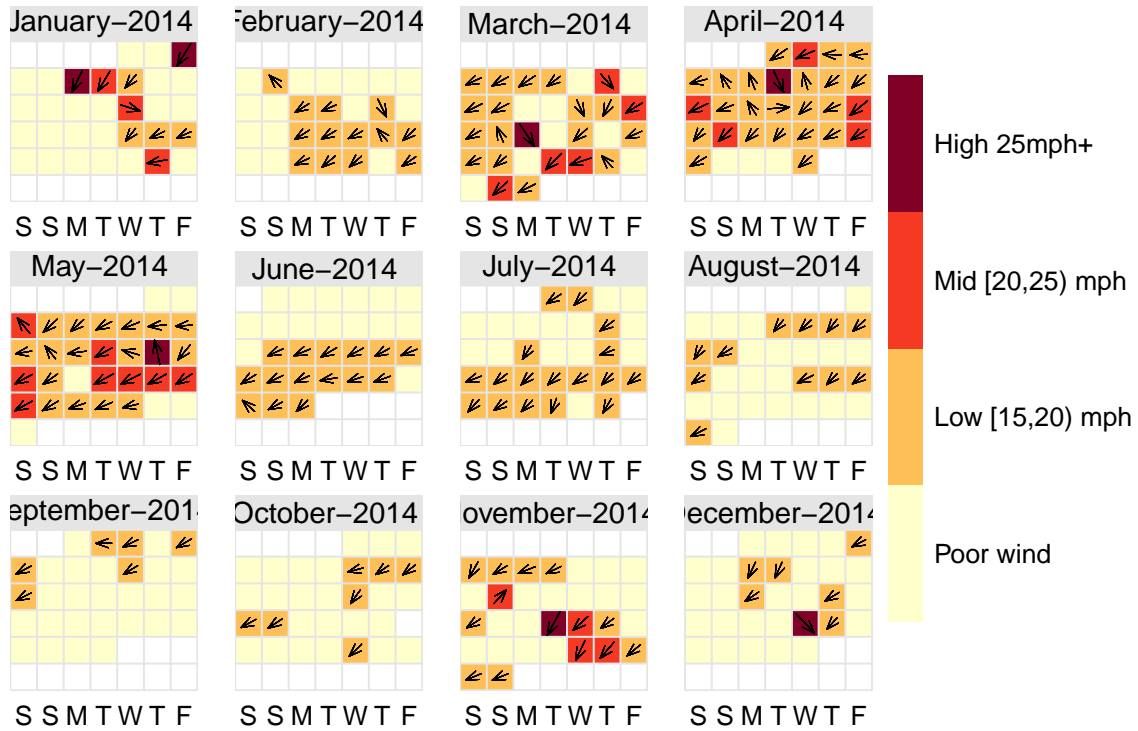
## B. Remaining calendar plots

*Note* : No color on a day represents a missing record.

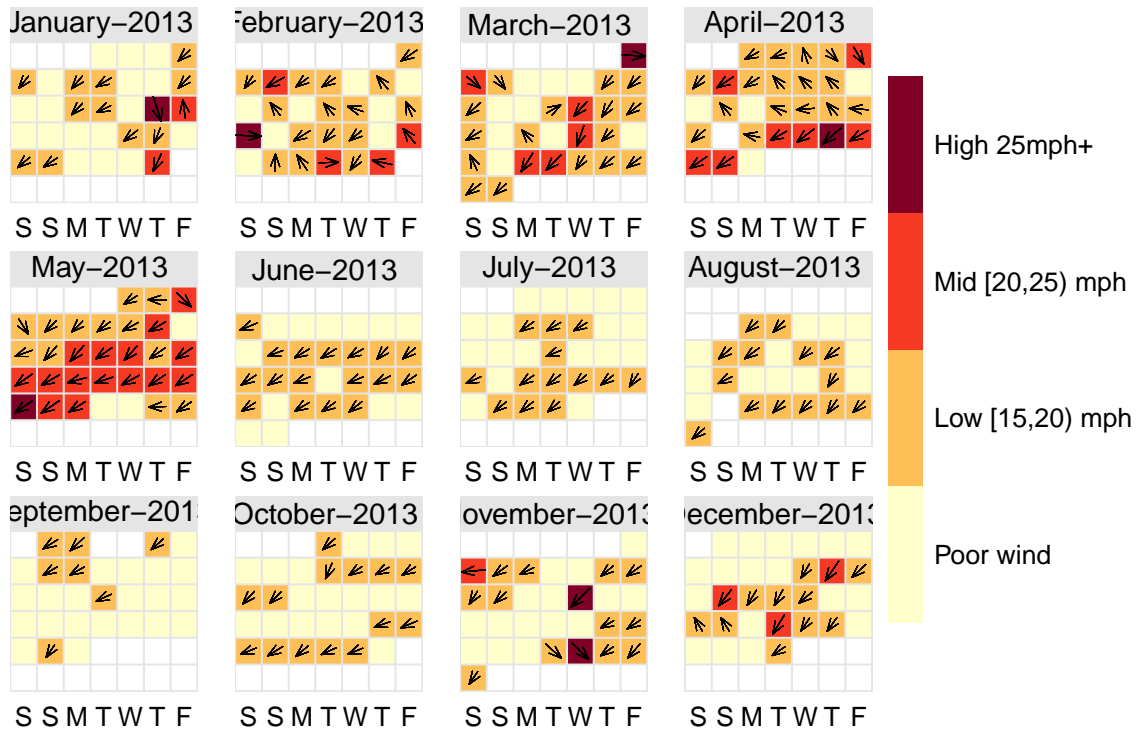
### Kiting days in 2015



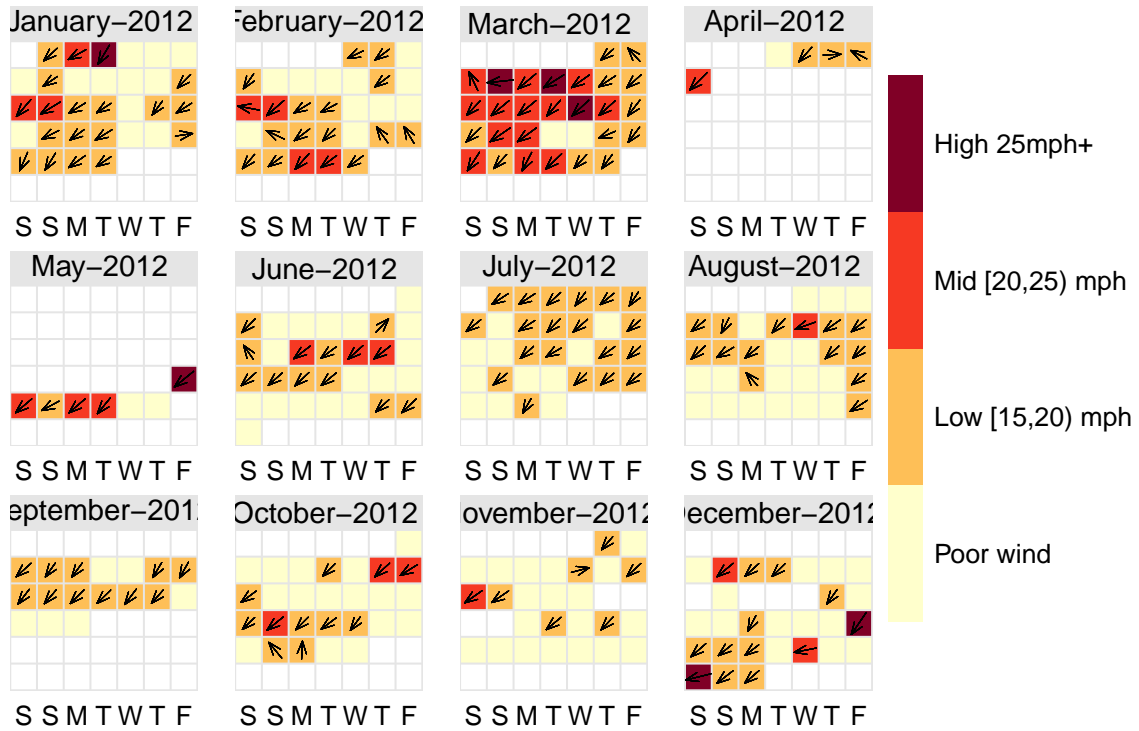
## Kiting days in 2014



## Kiting days in 2013



## Kiting days in 2012



### C. Runtime

It took the following amount of time to generate this report:

## Time difference of 4.990532 mins