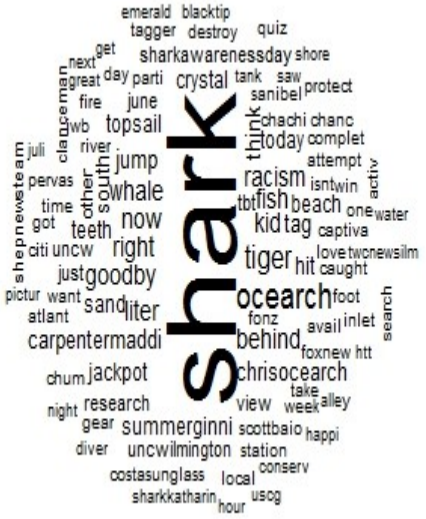
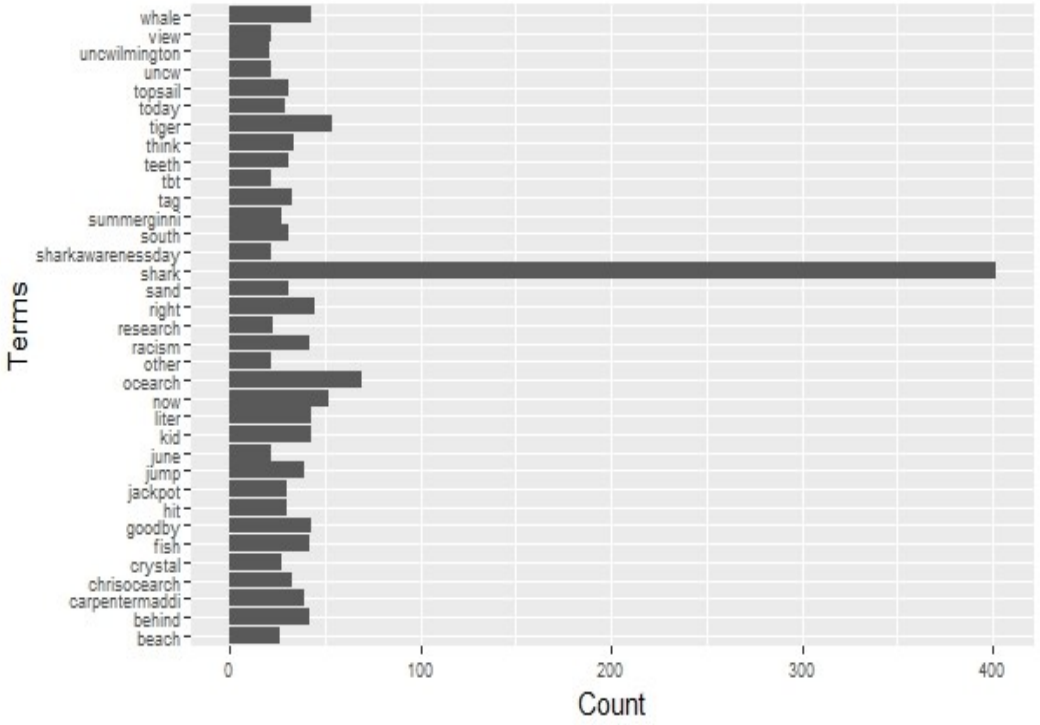
Twitter Mining on Shark Attacks

North Carolina and South Carolina

WORD CLOUD- NC

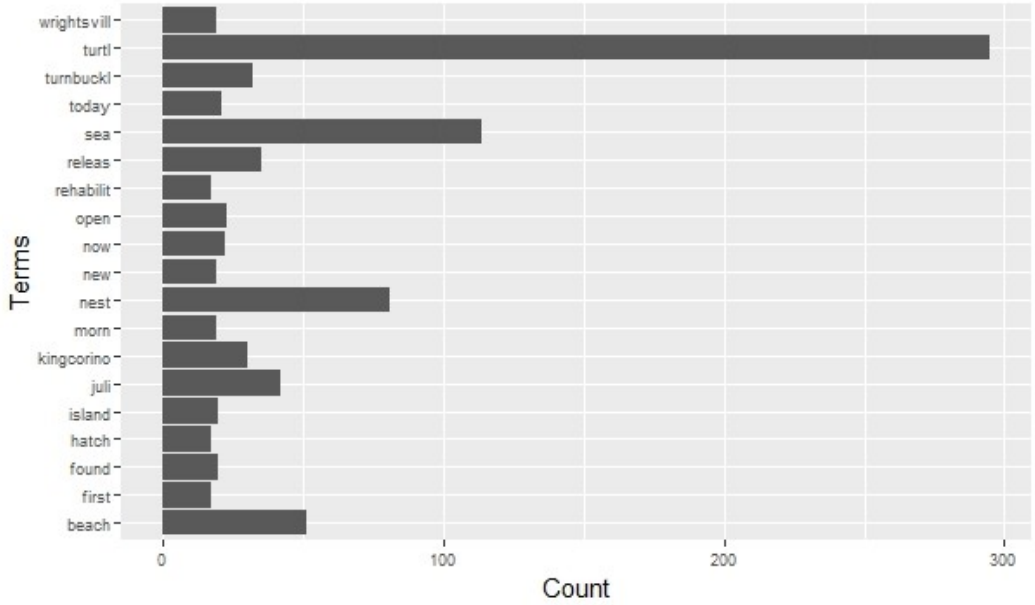
- Word Cloud for the keyword “Shark OR #Shark”





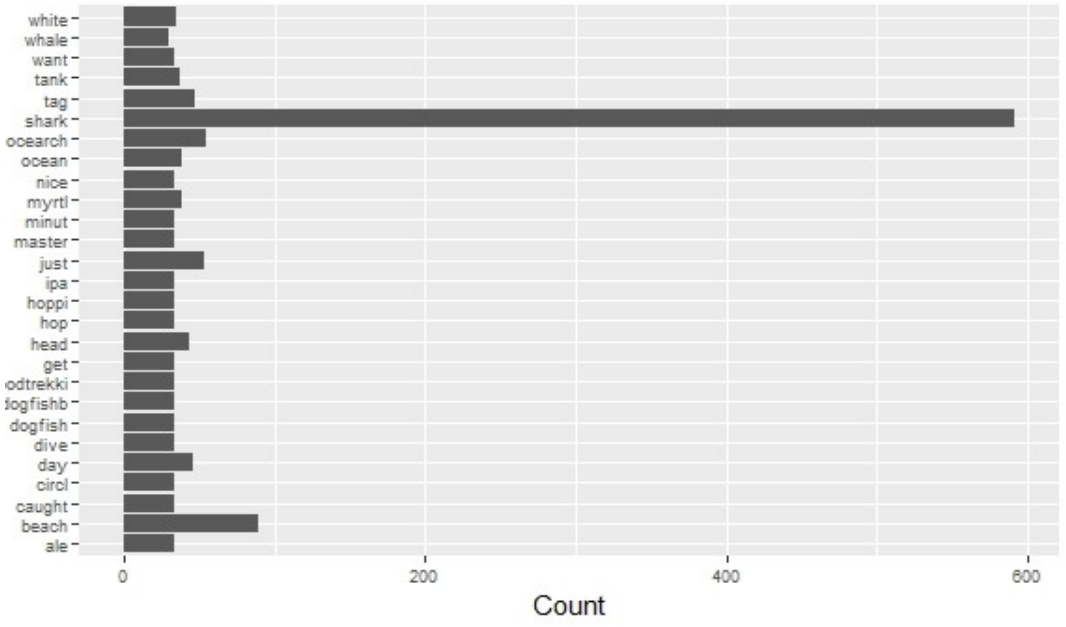
…Continued

- Word Cloud for the keyword “Turtle OR #Turtle”



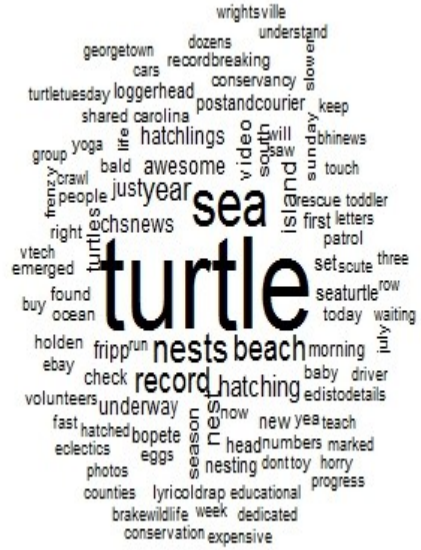
WORD CLOUD- SC

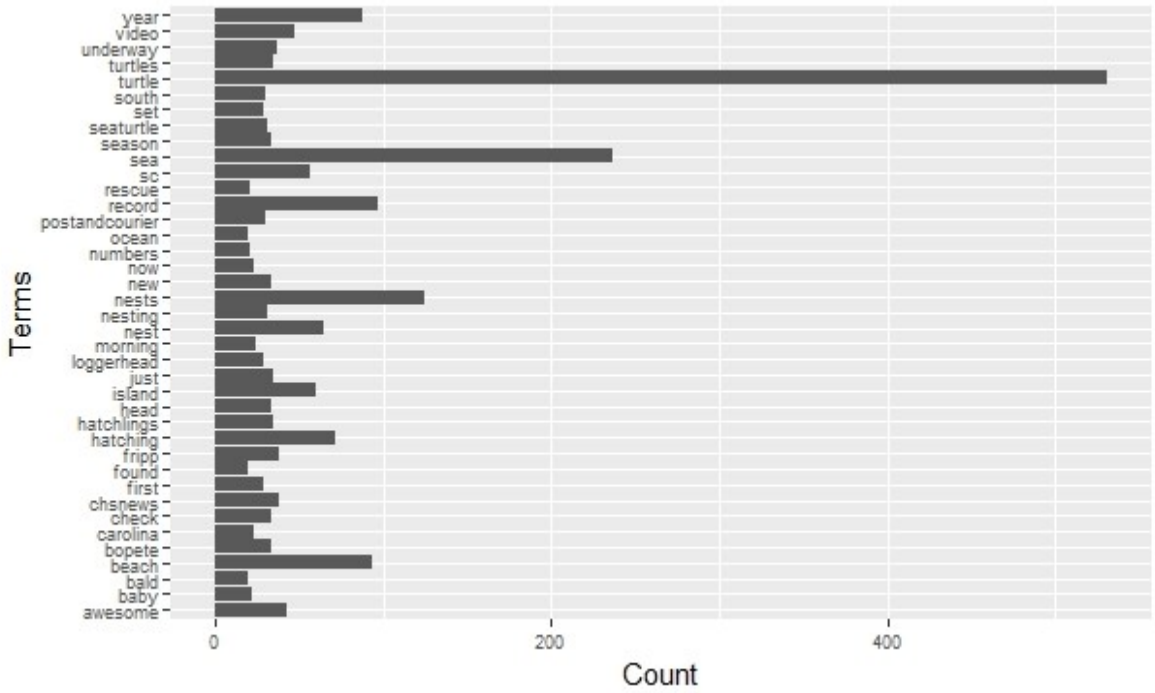
- Word Cloud for the keyword “Shark OR #Shark”



…Continued

- Word Cloud for the keyword “Turtle OR #Turtle”





RDATAMINING

1. Twitter Authentication

```
Console C:/Users/saileshbh/Desktop/Social mining with R/Social mining with R/
> library(devtools)
> library(twitteR)
> Consumer_key<- "YOUR API KEY"
> Consumer_secret <- "YOUR API SECRET"
> access_token <- "YOUR ACCESS TOKEN"
> access_token_secret <- "YOUR ACCESS TOKEN SECRET"
> setup_twitter_oauth(Consumer_key,Consumer_secret,access_token,access_token_secret)
```

2. Extracting Tweets

NCsharkData<- searchTwitter("Shark OR #Shark", n=1000, lang="en", geocode="35.7596,-79.0193,50mi", since="2009-05-01"))

…Continued

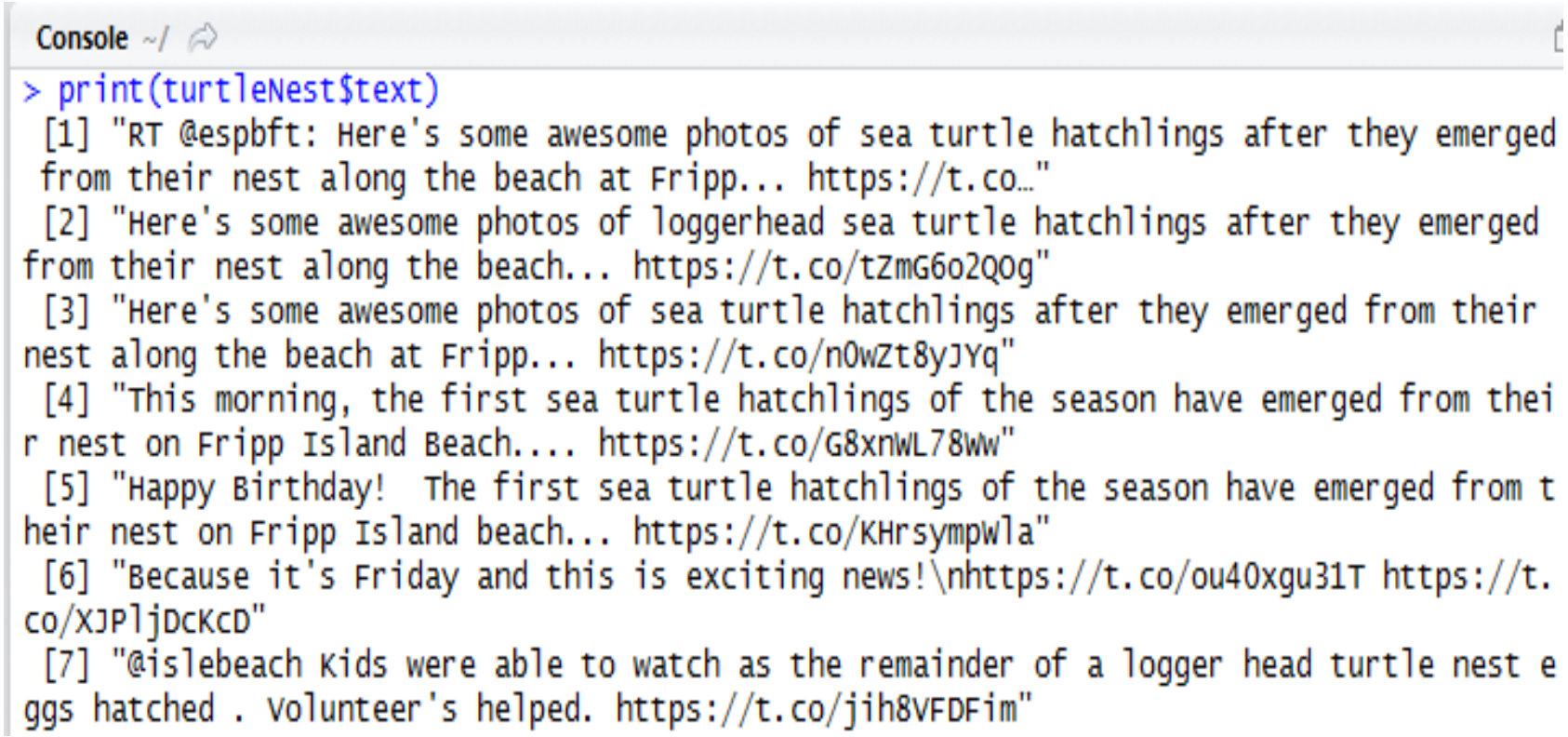
3.Text Cleaning

```
>ncs <- twListToDF(NCSharkData)
>writeLines(strwrap(ncs$text,60))
>myCorpus <- Corpus(VectorSource(ncs$text))
>removeURL <- function(x) gsub("http[^[:space:]]*", "", x)
>myCorpus <- tm_map(myCorpus, content_transformer(removeURL))
>removeNumPunct <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)
>myCorpus <- tm_map(myCorpus, content_transformer(removeNumPunct))
>myCorpus <- tm_map(myCorpus, content_transformer(tolower))
>myStopwords <- c(setdiff(stopwords('english'), c("r", "big")),
                 "use", "see", "used", "via", "amp")
>myCorpus <- tm_map(myCorpus, removeWords, myStopwords)
>myCorpus <- tm_map(myCorpus, stripWhitespace)
>myCorpusCopy <- myCorpus
>myCorpus <- tm_map(myCorpus, stemDocument)
```

4. Word Cloud

wordcloud(myCorpus, max.words = 100, random.order = FALSE)

SAMPLE TWEETS



LIMITATIONS OF TWITTER ACCESS

- **Search**

  Rate limiting of the API is primarily considered on a per-user basis

  Search is rate limited at 180 queries per 15 minute window.

  When an application exceeds the rate limit for a given API endpoint, the

  Twitter API will now return an HTTP 429 “Too Many Requests” response code.

MINING OTHER SOCIAL NETWORKS

Facebook:

- Like Twitter, First we have to create an App on the Facebook platform. We will use this app to connect to the Facebook API.
- Install the packages devtools and Rfacebook in “R”
- The getUsers function returns public information about one or more Facebook user. If we use “me” as the username argument, it will return our own profile info.

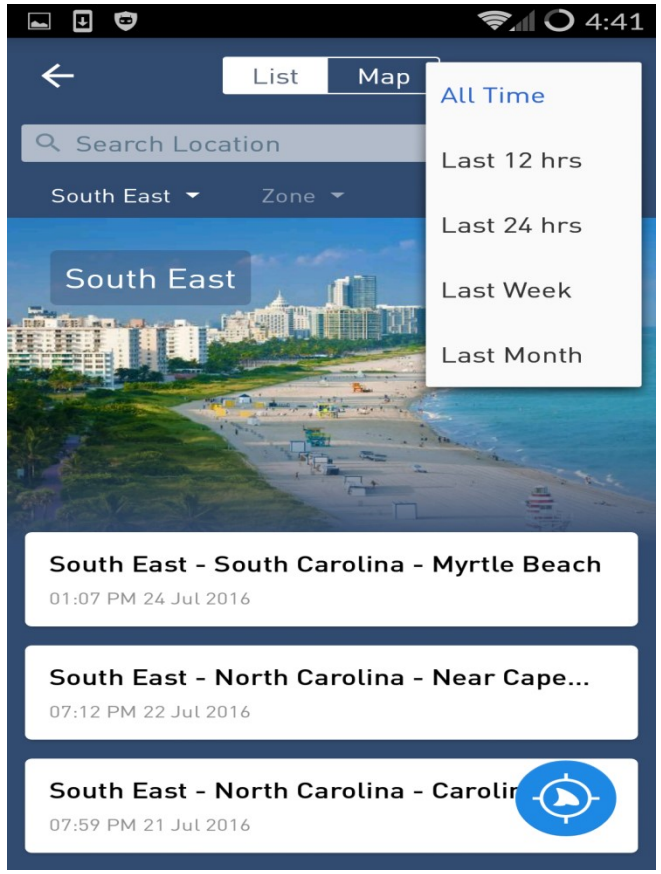
Instagram:

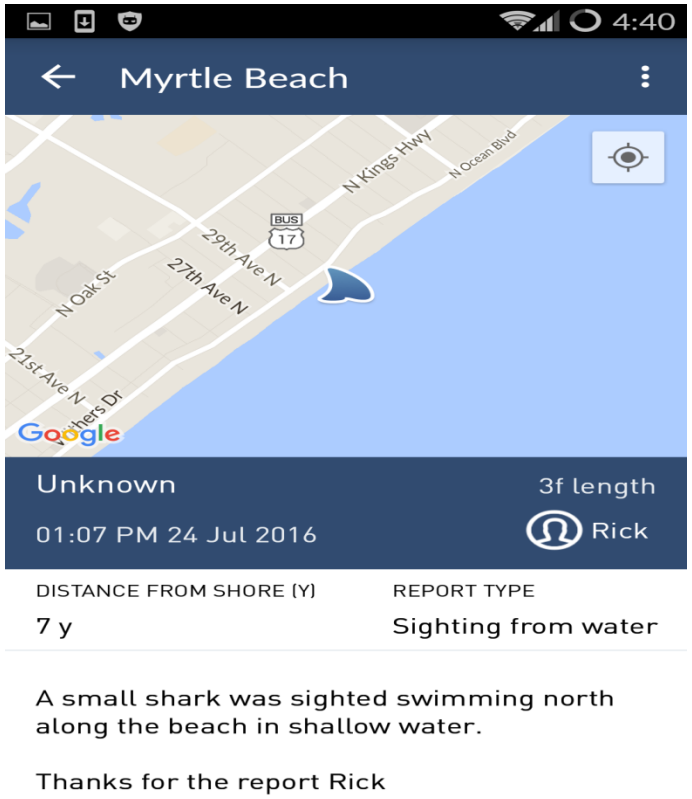
- There is no R package for this yet so we have to configure the authentication and data download process on our own
- Instagram uses oAuth 2 which makes it easy to use with R and the httr package for example.
- we create our Instagram in R for the httr package. This is the app we will use to connect to the API

```
instagram <- oauth_endpoint(
  authorize = "https://api.instagram.com/oauth/authorize",
  access = "https://api.instagram.com/oauth/access_token")
myapp <- oauth_app(app_name, client_id, client_secret)
```

MORE RESOURCES

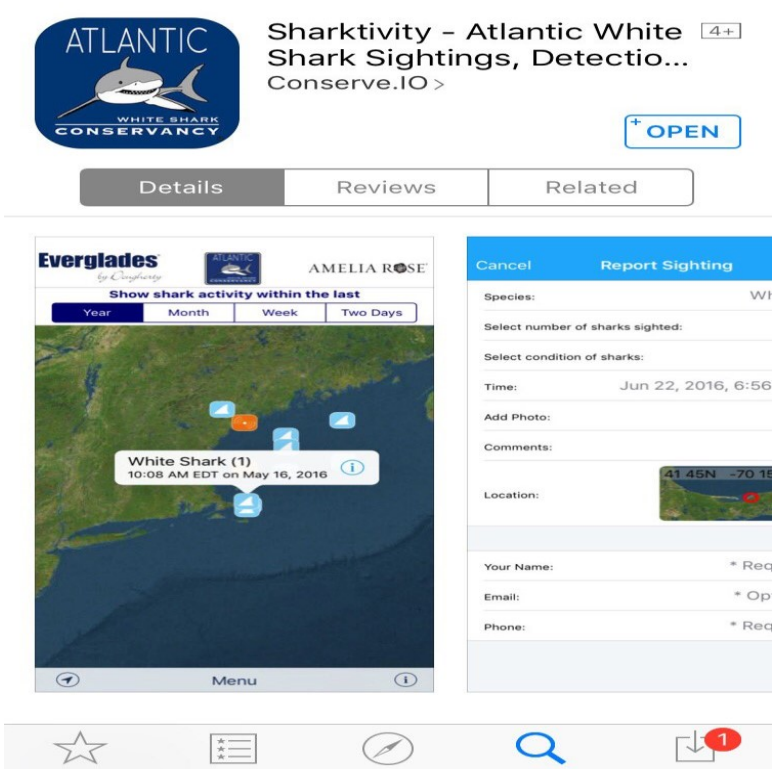
- **Dorsal:** The world's most advanced shark reporting and alerts solution. It is a free community based shark alert solution that allows beachgoers and authorities to immediately alert others to shark sightings or attacks in their area – they use #sharkreport on Twitter

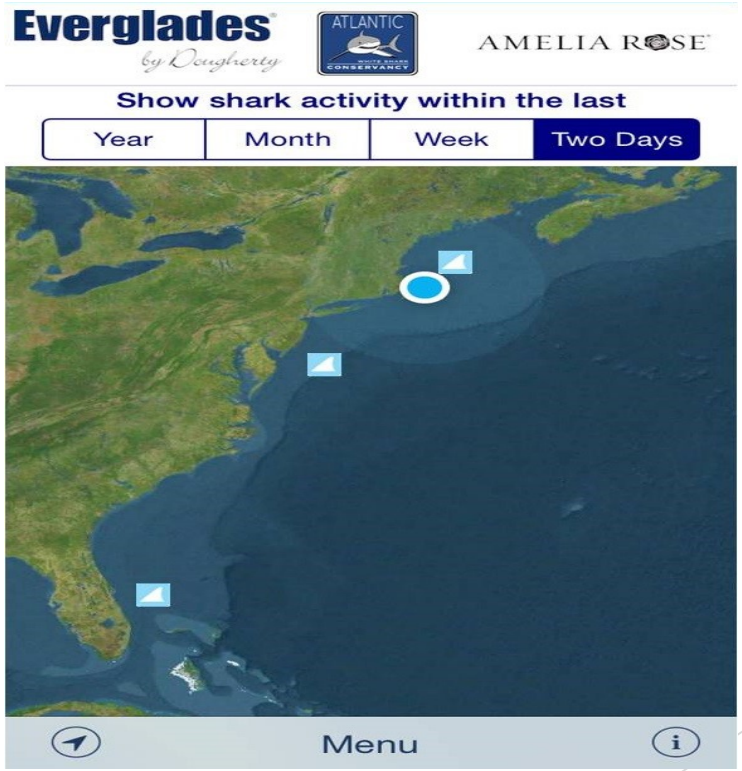




…Continued

- **Sharktivity:** The application lets users report their own shark encounters, track the movements of previously tagged apex predators, and view recent visual sightings by experts.
- "This app will be yet another tool to raise awareness and provide the public with information on white shark activity across the East Coast," conservancy president Cynthia Wigren said in a statement.





…Continued

- **Ocearch:** Ocearch is a recognized world leader in generating critical scientific data related to tracking (telemetry) and biological studies of keystone marine species such as great white sharks, in conjunction with conservation outreach and education at a measurable global scale.
- Ocearch shares real-time data through OCEARCH's Global Shark Tracker, inspires current and future generations of explorers, scientists, and stewards of the ocean, and enables leading researchers and institutions to generate previously unattainable data.
- Ocearch has completed 22 expeditions as of September 2015; by 2016, a total of 26 will be completed.

HASHTAG STANDARDIZATION

- Hashtags allow researchers, emergency responders and affected community members to easily identify twitter conversations related to a given topic.

- **Challenges:**

  a. No uniformity in the system

  b. Large number of similarly written retweets may become confusing, especially when it comes to verifying the original author and validity of the message.

- **Proposed standards:**

  a. Encourage geo-location of tweets

  b. Use of Infographics: The use of Infographics allow Twitter users to easily retweet the important information without the 140-character limitation.

…Continued

Conclusion and Recommendations:

- Although there are many limitations, we also acknowledge that no monitoring system will be perfect in terms of low-cost, real-time analysis and high accuracy.
- Alternative to GPS location,What3Words can also be used. This program exchanges longitude and latitude lines for three words to help communicate location information.
- Easy retrieval of Information: Hashtag must be completely specific, should not mimic a country's emergency phone system (e.g., #911US)
- Grouping of hashtags: unique hashtag must be used for all the tweets conveying similar information. This will help emergency responders to identify the topic easily