A Step by Step How To for Extracting Twitter Messages from R

I recently started a small hobby project to analyse accident frequency on Singapore roads. I decided to extract this information from the Singapore Land Transport Authority twitter feed. (although I could have gotten data through the [DataMall](http://www.mytransport.sg/content/mytransport/home/dataMall.html) initiative by the Singapore Government using Python, this would be the subject of another how-to later ☺ )

I thought I would share my experience and steps to do this. So what are we waiting for? Let’s start!

**Step 1: Download the twitteR package**

We need to ensure that the latest twitteR package is installed on your R environment. Run the following command in R Studio

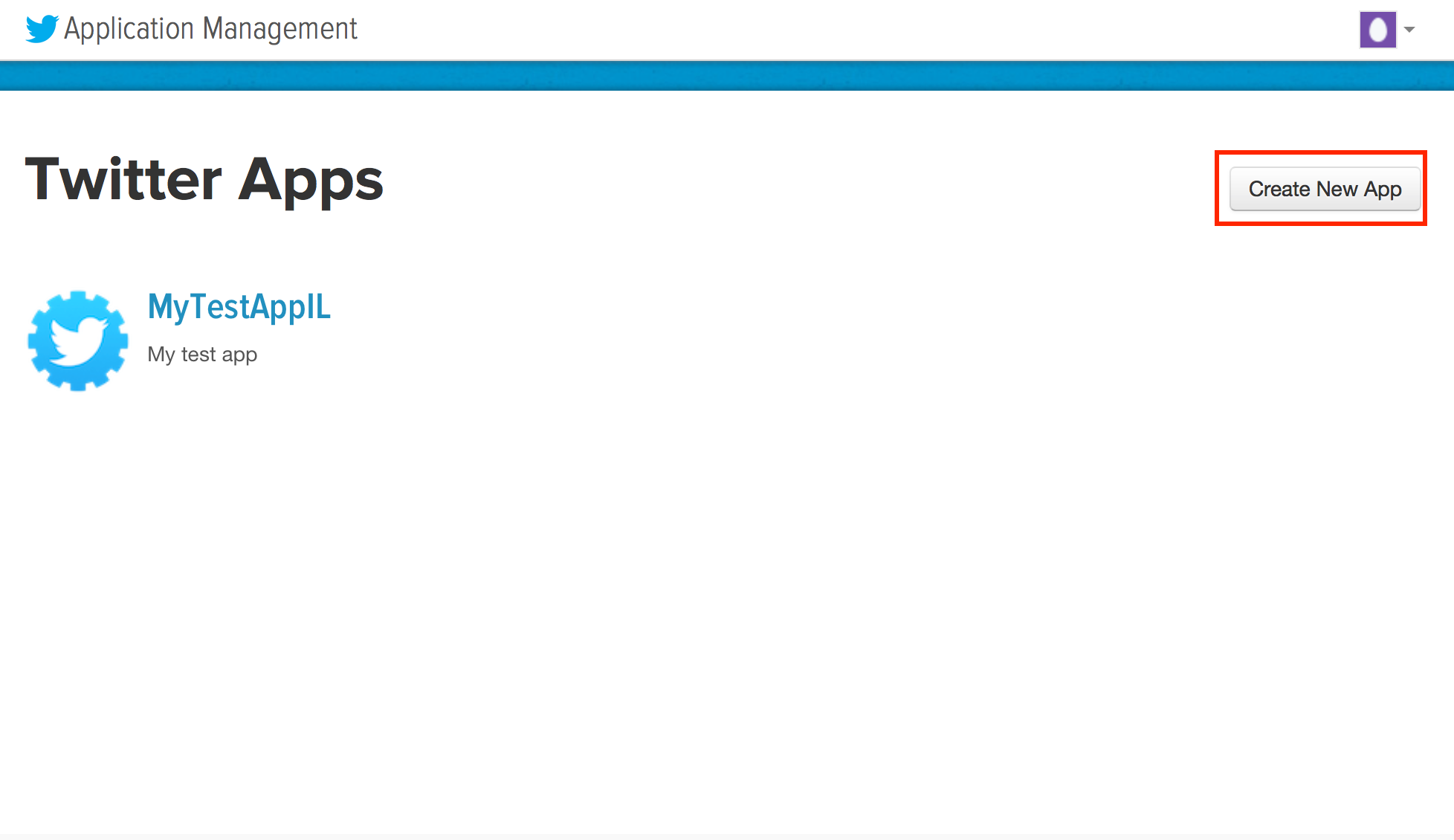
install.packages (twitteR)

This will download and install the twitteR and all required packages.

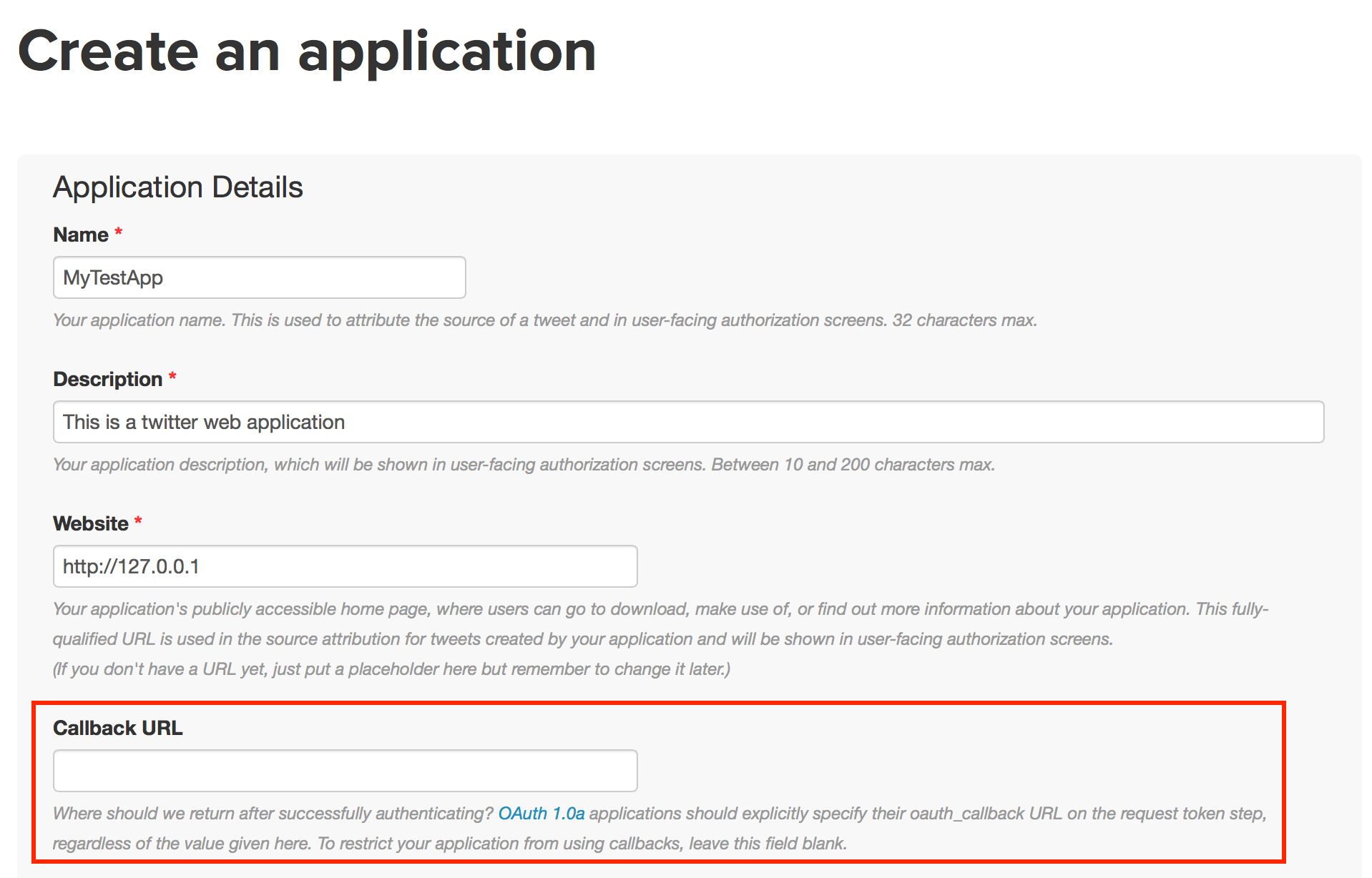
**Step 2: Setup a Twitter App**

We need to create a Twitter App so that we can access the Twitter platform through this web API. Before you can create a Twitter App, you need to create an account first. You can do so on the [Twitter Apps](https://apps.twitter.com/) page.

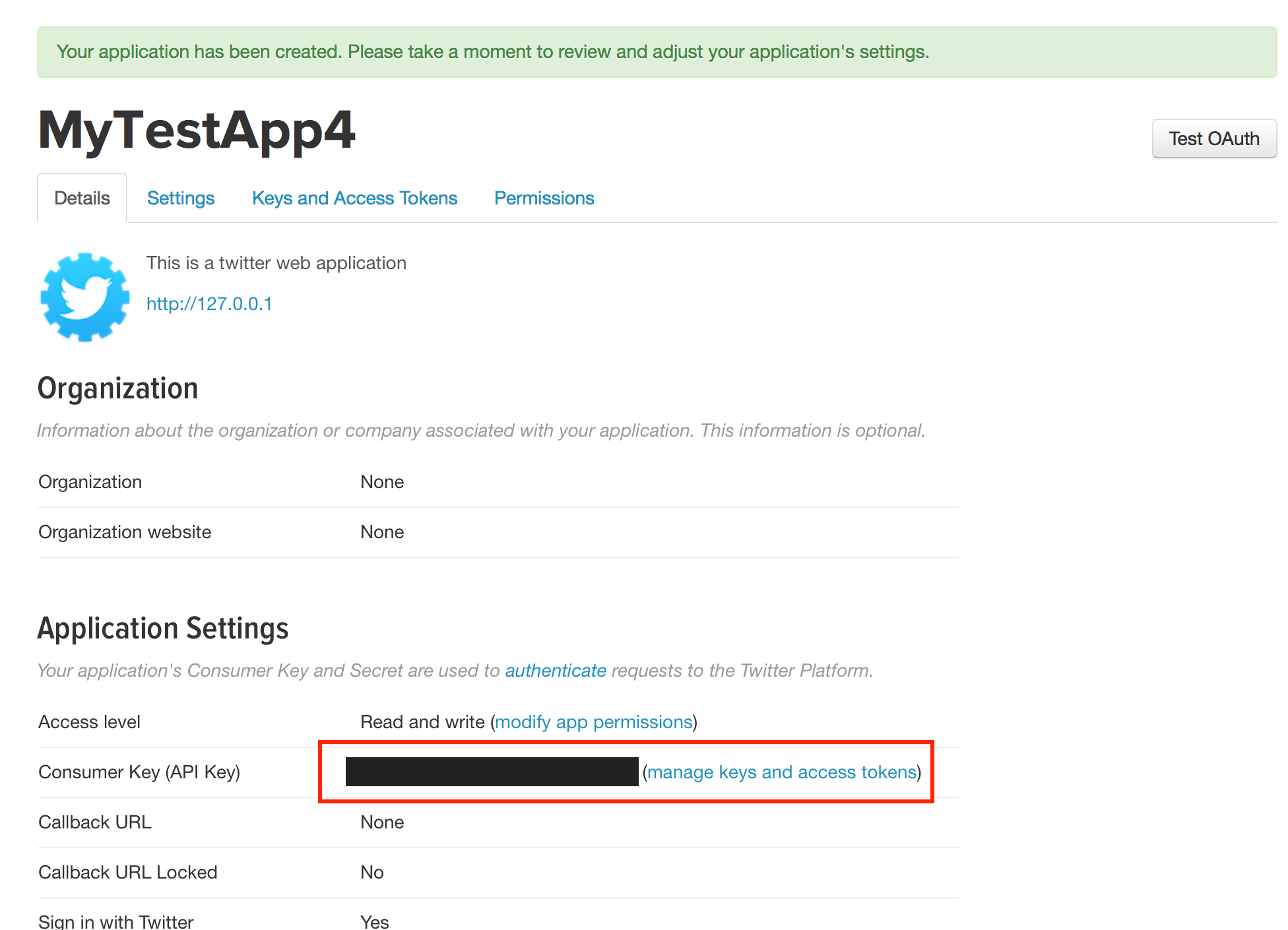
Once you are done, you can start by clicking on the Create New App button.



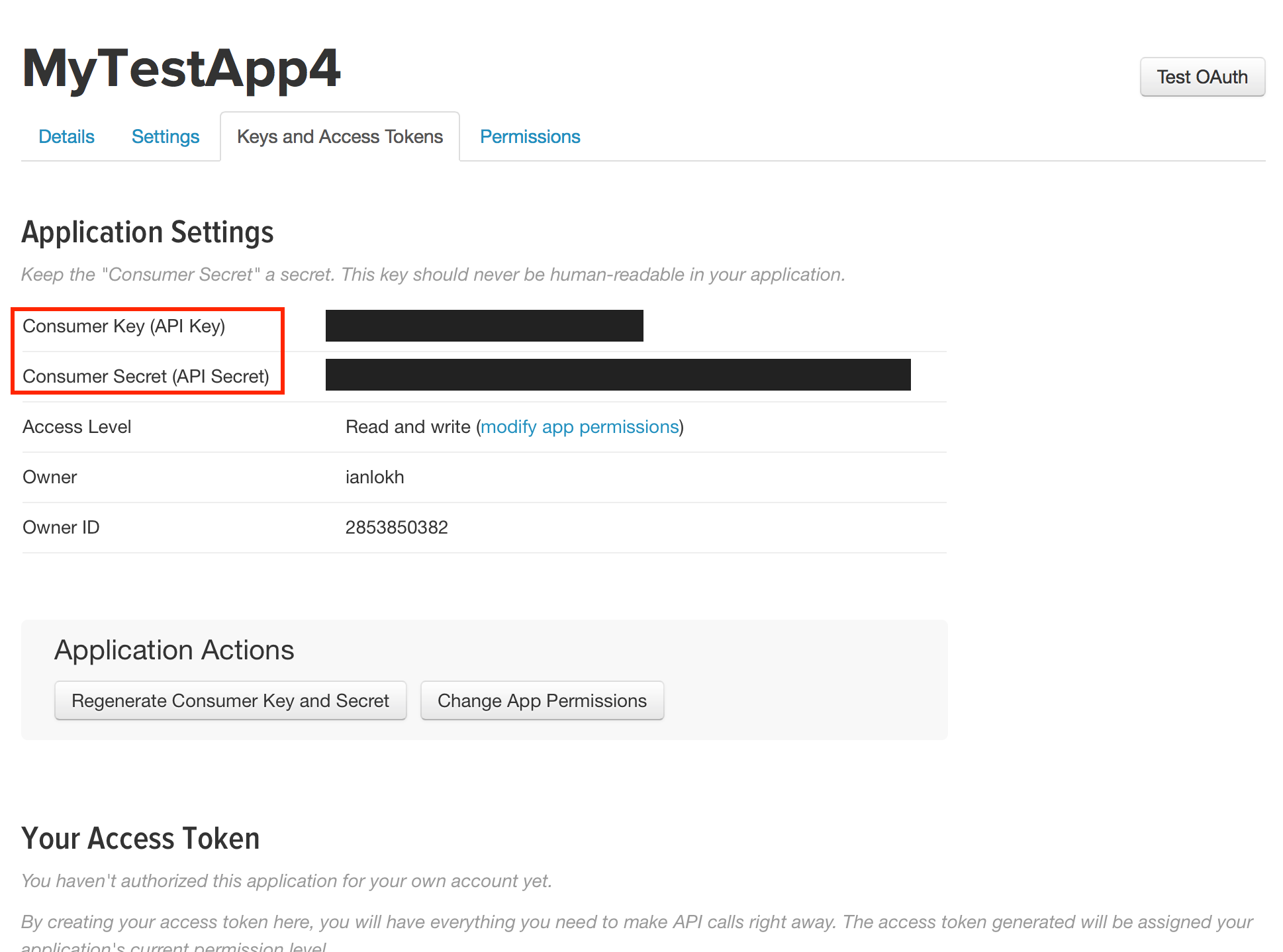
Proceed to enter the required mandatory fields as shown below.



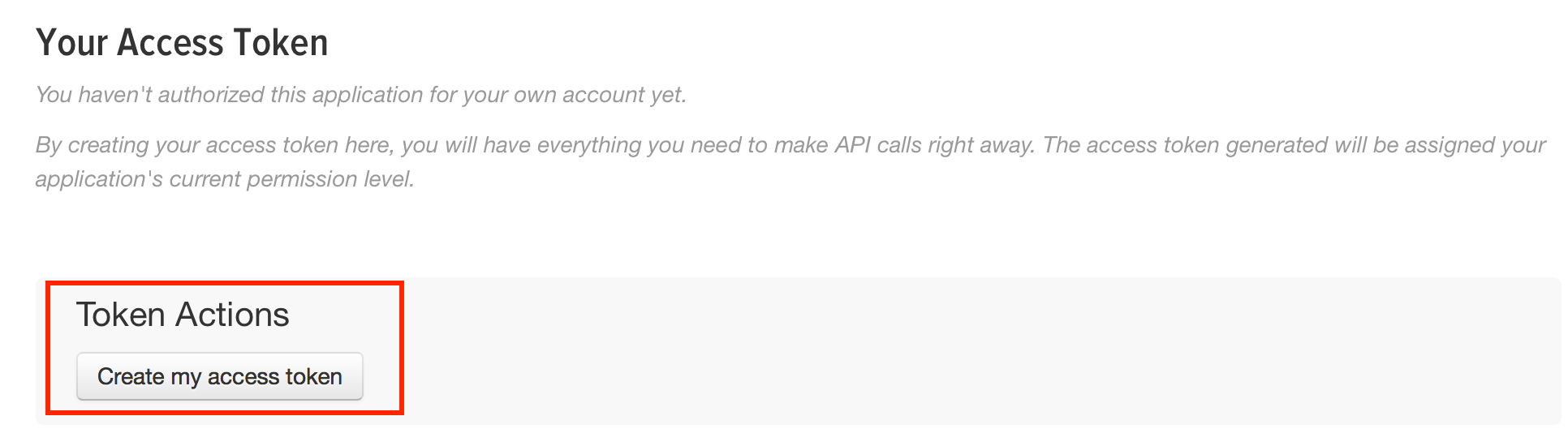
The Website address can be a temporary one for now. However, ensure that the Callback URL is left blank for now.

Acknowledge the developer agreement and click on the “Create your Twitter application” button. The following page will appear confirming that you have successfully create the web application. 

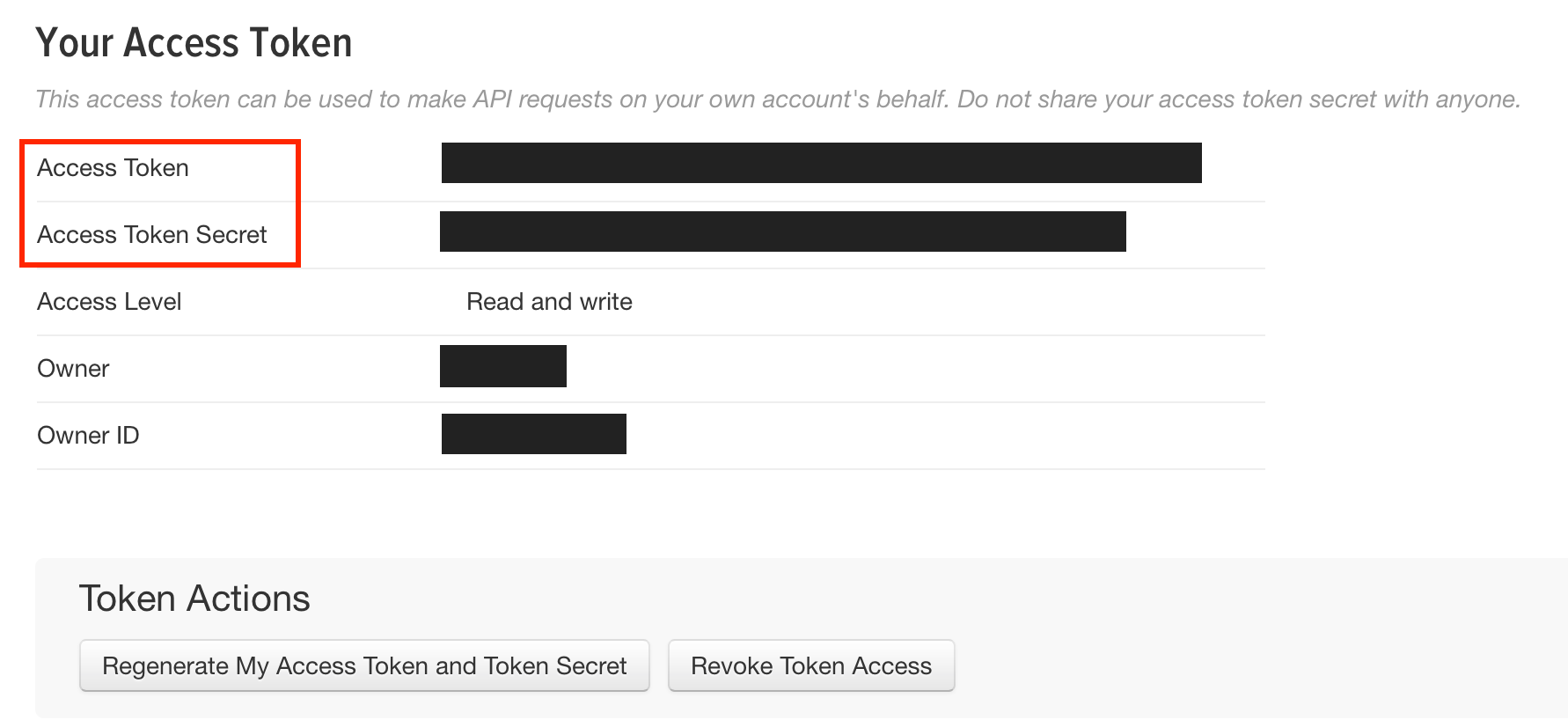
Click on the Keys and Access Token tab to view the Consumer Key and Consumer Secret keys.



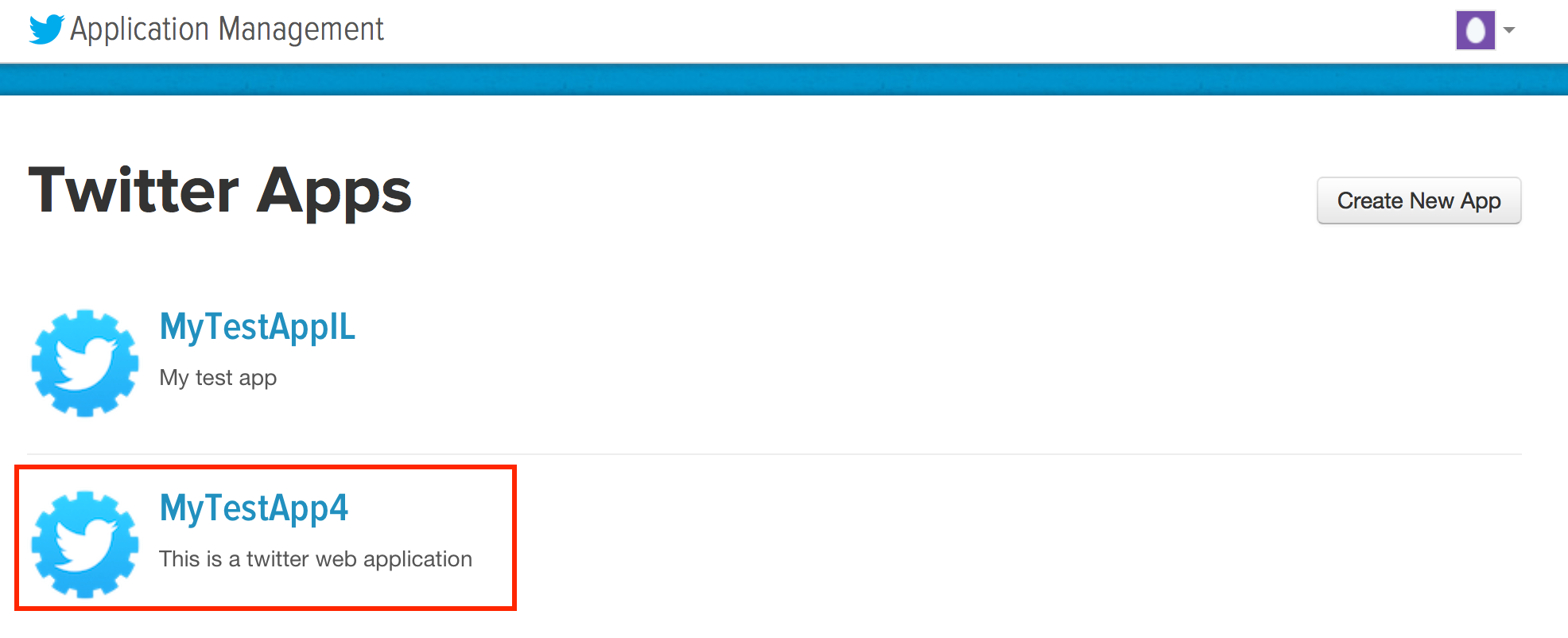
At this point, you have not created your Access Token yet. Hence click on “Create my access token” button to do so.



Your access tokens will be generated and displayed on the refreshed page.



Click on the Application Management icon above and you will see your new application created as shown below.



**Step 3: Create R code to Access Twitter Feeds**

Go back to RStudio and enter the following R code:

#install the necessary packages

library(twitteR)

#necessary file for Windows

#download.file(url="http://curl.haxx.se/ca/cacert.pem", destfile="cacert.pem")

#to get your consumerKey and consumerSecret see the twitteR documentation for instructions

consumer\_key <- 'your consumer key'

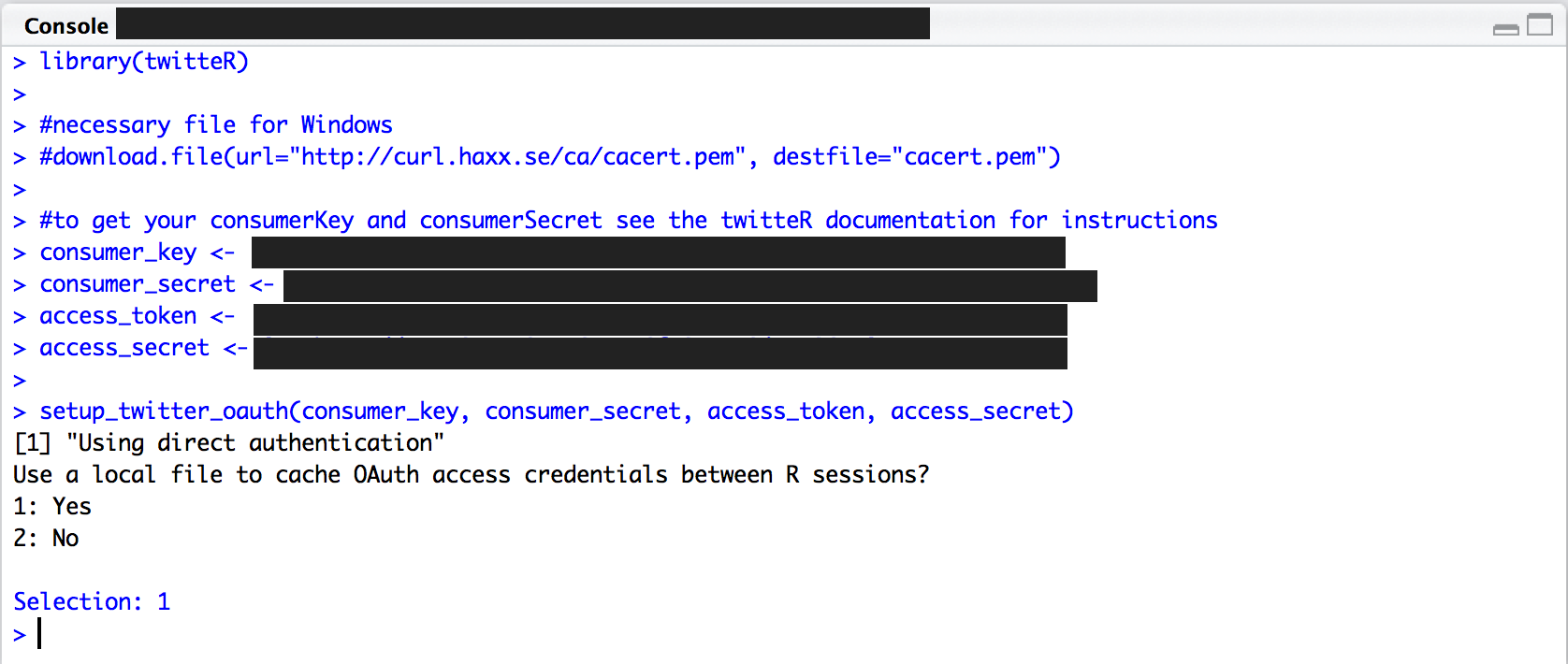
consumer\_secret <- 'your consumer secret key'

access\_token <- ‘your access token’

access\_secret <- 'your access secret’

setup\_twitter\_oauth(consumer\_key, consumer\_secret, access\_token, access\_secret)

Once you have entered the above, run the code and you will see the following prompt on the RStudio console



You can select 1 or 2 depending on your preference. Regardless of the choice, you should see the “>” on the next line on the console indicating that the setup\_twitter\_oauth command was successfully executed.

**Step 4: Extract your Twitter Feed**

Once you have completed the above step, enter the following R code.

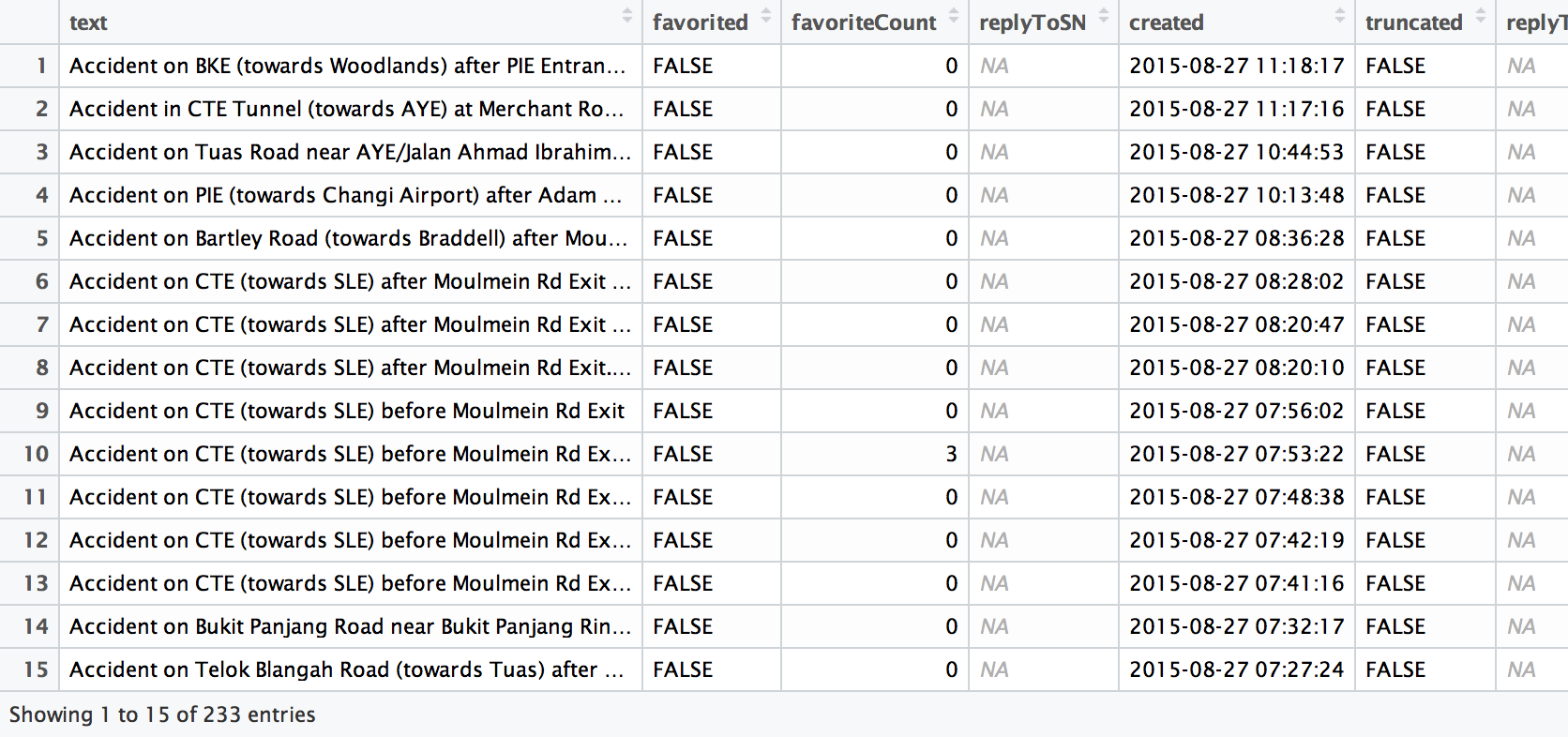
ltaTwtr <- searchTwitter("<string that you want to extract from>", n=500)

length(ltaTwtr)

#make data frame

tmpDf <- do.call("rbind", lapply(ltaTwtr, as.data.frame))

The command searchTwitter will issue a search of Twitter based on a supplied search string – based on your subscribed twitter feeds. Because the return value of searchTwitter is a list, we would need to do.call(“rbind”…) function to convert it into a data frame for subsequent processing.



The above table is an example of the twitter messages that match my search criterion.

Hope this short how-to has help with your data science tasks!