Twitter Mining in R

Dr Akhter Raza

Installing packages

```
install.packages("twitteR")
install.packages("RCurl")

require("twitteR")
require("RCurl")
```

Getting twitter keys

create a twitter account and go to the url apps.twitter.com and collect four keys from there

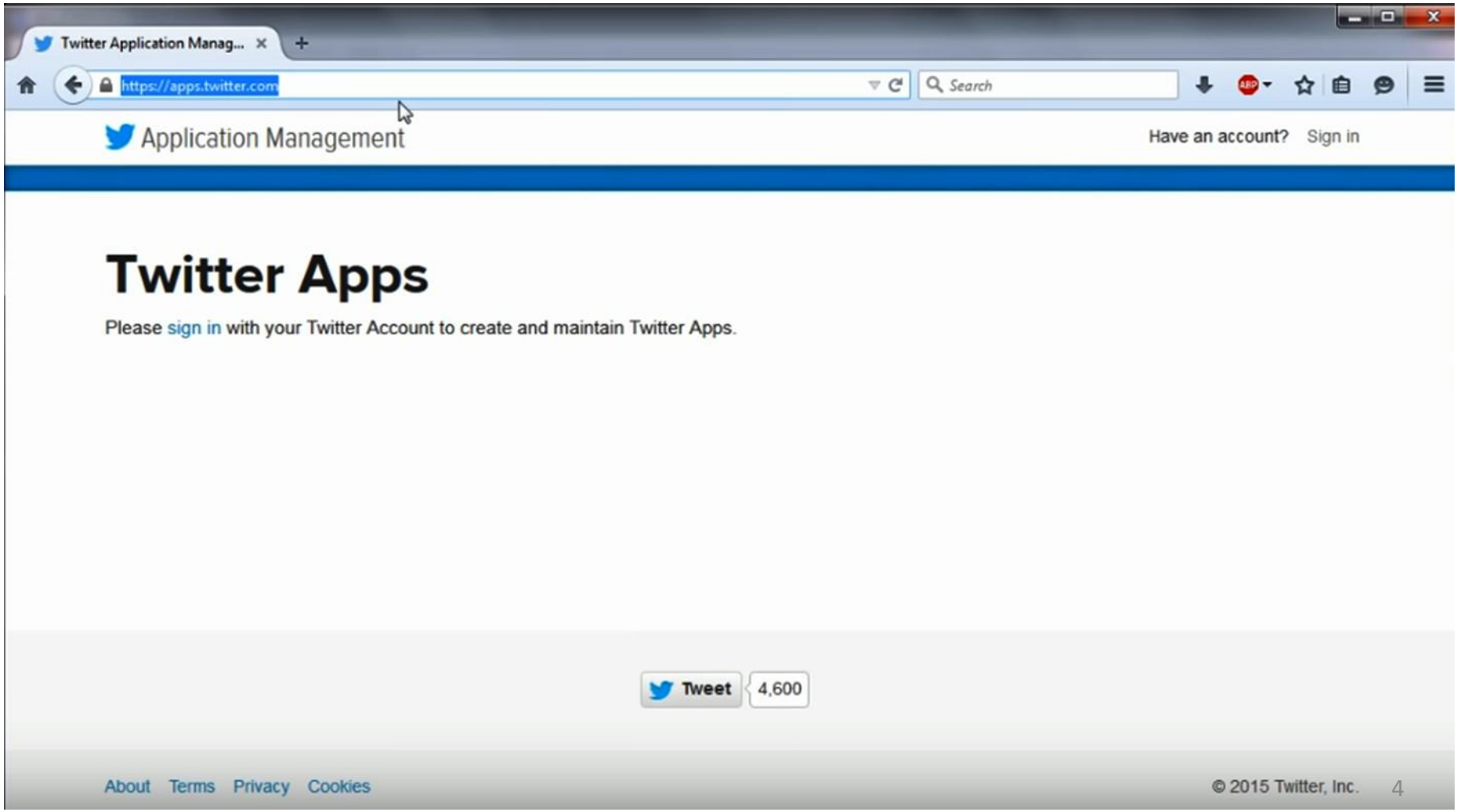
consumer_key

consumer_Secret

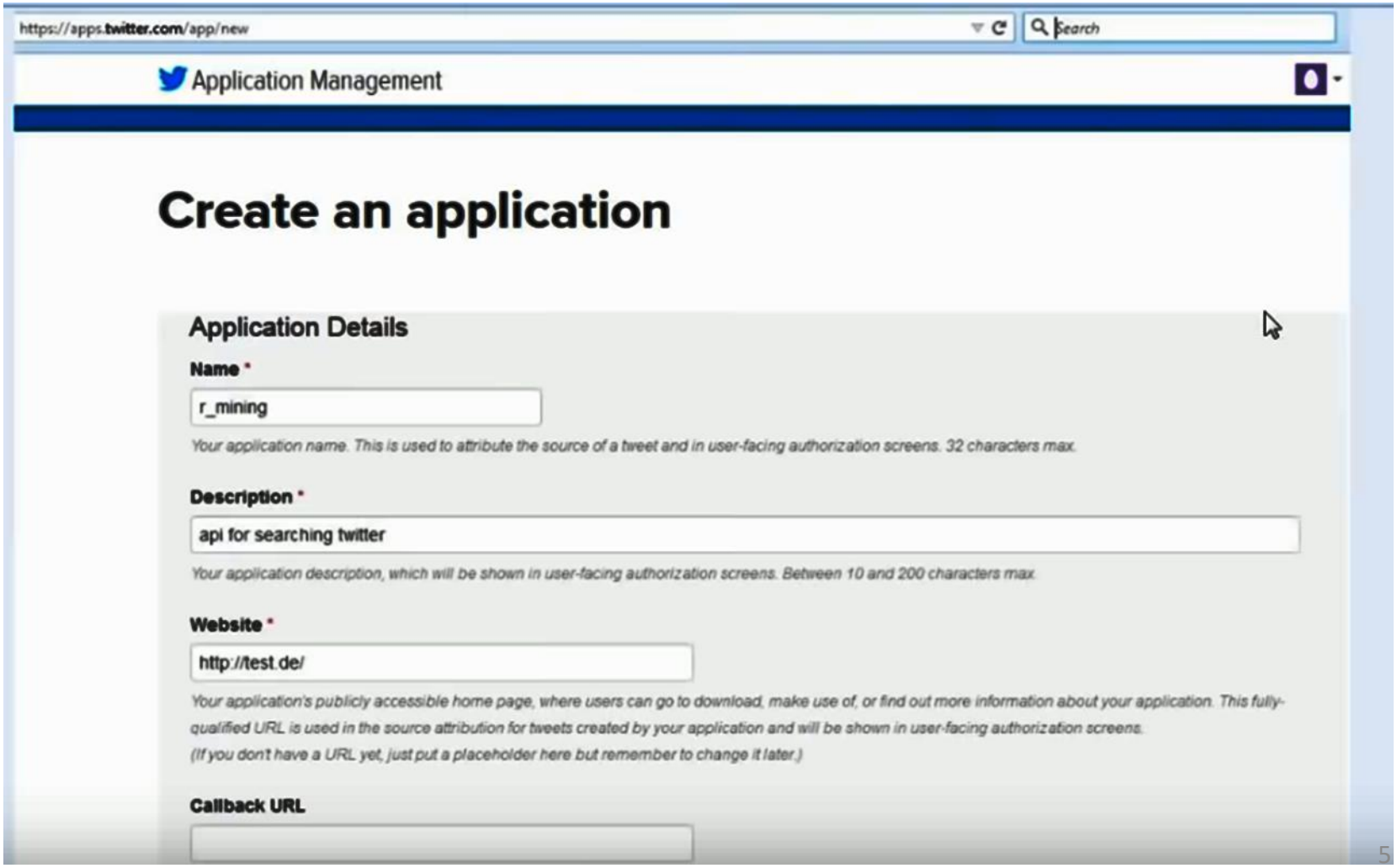
access_token

access_secret

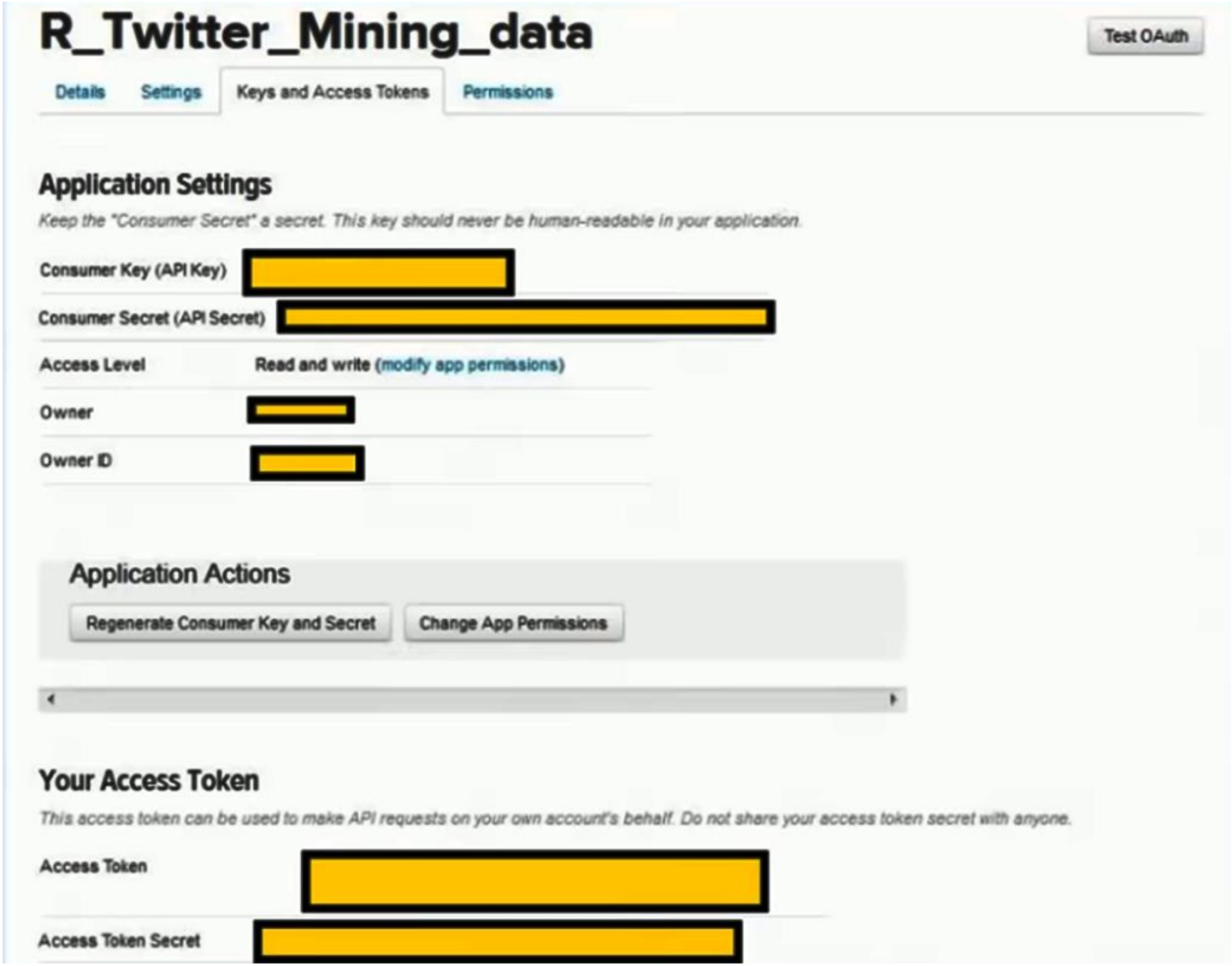
Creating twitter App



Creating twitter App



Creating twitter App



Setting up twitter keys

```
consumer_key      <-   ""

consumer_Secret   <-   ""

access_token      <-   ""

access_secret     <-   ""
```

Authentication & Searching

```
setup_twitter_oauth(consumer_key,consumer_Secret,access_token,access_secret)
```

Now search for liverpool football club

```
LFC_tweets <-  searchTwitter("LFC", n=10, lang = "en")
```

Showing tweets

LFC_tweets

str(LFC_tweets)

LFC_tweets[1:3]

Showing LFC tweets

First three tweets from LFC

```
> LFC_tweets[1:3]
[[1]]
[1] "NizamuddinAwang: RT @MPBFirmino9: Few weeks back United and Spurs
 fans were laughing that they would derail #LFC's title challenge and
save football by help…"

[[2]]
[1] "RhysAshcroft1: I blame @LFC for being so involved in my dreams la
st night that im now suuuper late for work"

[[3]]
[1] "junior4run: Liverpool reserve team have more spectators than Manc
ity vs Cardiff last night. #YNWA #lfc"
```

Text mining of tweets

Two packages tm and wordcloud are needed

```
install.packages("tm")
install.packages("wordcloud")
require(tm)
require(wordcloud)
```

Searching tweets

now searching tweets for earthquake

```
himearth <-
searchTwitter('earthquake+himalaya', lang =
en", n= 50 , resultType = "recent")

class(himearth)
```

Converting list into chr

himearth is a list we convert into a character vector

```
himearth_text <- sapply(himearth, function(x)
x$getText())
```

```
str(himearth_text)
```

```
> str(himearth_text)
 chr [1:500] "Nepal Mountain Villages 'Completely Washed Away' By Quake http://t.co/v$
```

Making corpus

Corpus is the main structure of managing text

```
him_corpus<- Corpus(VectorSource(himearth_text))
him_corpus
```

See into the corpus

```
inspect(him_corpus)
inspect(him_corpus[1])
```

Removing punctuations

now we clean up the corpus by eliminating the unnecessary words like by, and, punctuations etc

removing punctuations

```
him_clean <-
tm_map(him_corpus,removePunctuation)
```

Uppercase to lowercase

transforming all uppercase letters to lowercase

```
him_clean <-
tm_map(him_clean,content_transformer(tolower))
```

Removing stopwords, digits, whitespaces

Removing stopwords like and is or etc

```
him_clean <-
tm_map(him_clean,removeWords,stopwords("english"))
```

remove numbers and digits

```
him_clean <- tm_map(him_clean,removeNumbers)
```

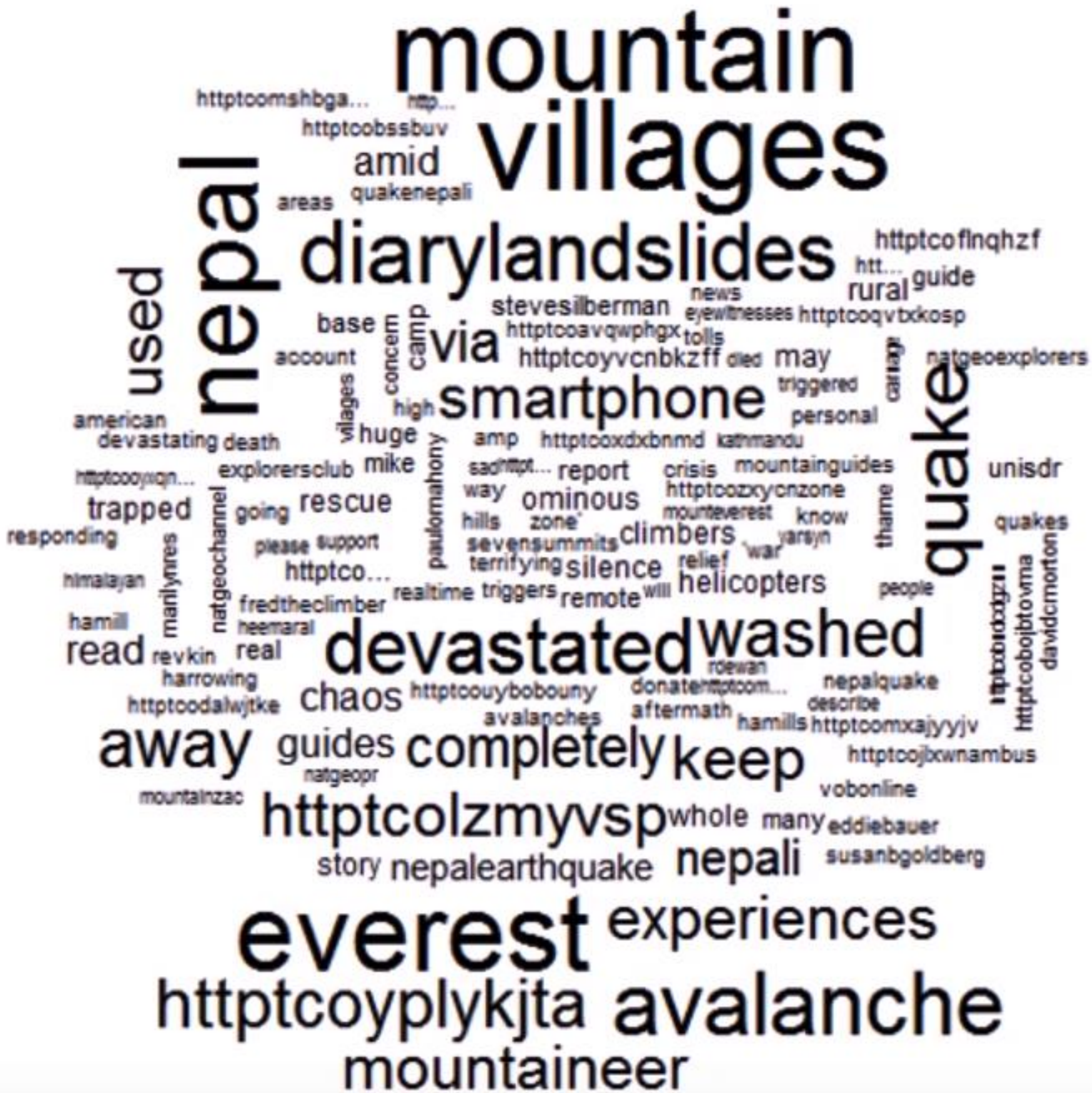
removing whitespace

```
him_clean <- tm_map(him_clean,stripWhitespace)
```

Removing any word

remove himalaya and earthquake b/c it is obvious

```
him_clean <- tm_map(him_clean,removeWords,
c("himalaya","earthquake"))
```


mountain
villages
nepal
diarylandslides
used
amid
areas
quakenepali
httptcoomshbga...
httptcoobssbuv
httptcoflnqhzf
guide
rural
news
stevesilberman
eyewitnesses
httptcoqvtxkosp
base
camp
concern
via
httptcoavqwphgx
tolls
httptcoyvcnbkzff
died
may
account
natgeoexplorers
triggered
smartphone
quake
high
villages
personal
american
devastating
death
huge
amp
httptcoxdxbnmd
kathmandu
mike
paulomahony
explorersclub
report
crisis
mountainguides
unisdr
way
ominous
httptcozxycnzone
trapped
going
rescue
hills
zone
mounteverest
know
thame
quakes
responding
please
support
sevensummits
climbers
'war'
marilynres
natgeochannel
terrifying
silence
relief
httptco...
helicopters
himalayan
realtime
triggers
remote
will
people
fredtheclimber
davidcmorton
hamill
heemaral
devastated
washed
read
revkin
real
harrowing
nepalquake
describe
httptcodalwjtke
chaos
httptcouybobouny
aftermath
avalanches
hamills
httptcomxajyyjv
away
guides
completely
keep
httptcojbxwnambus
natgeopr
vobonline
mountainzac
httptcolzmyvsp
whole
many
eddiebauer
story
nepalearthquake
nepali
susanbgoldberg
everest
experiences
httptcoyplykjta
avalanche
mountaineer

Graph of words by wordcloud

wordcloud(him_clean)

wordcloud(him_clean,random.order = F)

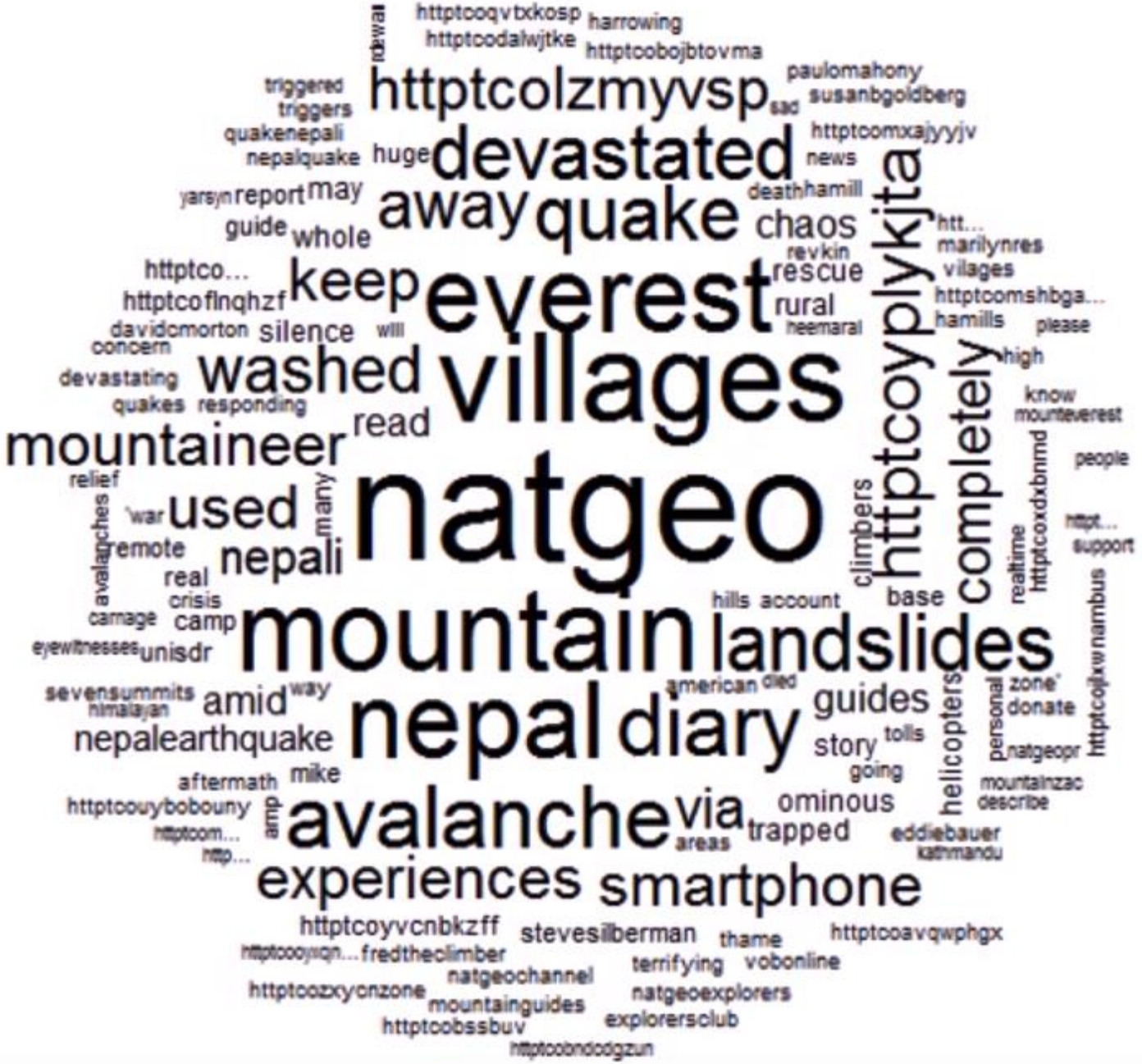
defining scale to enlarge the words min and max

wordcloud(him_clean,random.order = F, scale = c(3,.5))

Cleaning commands

```
> him_clean <- tm_map(him_corpus, removePunctuation)
> him_clean <- tm_map(him_clean, content_transformer(tolower))
> him_clean <- tm_map(him_clean, removeWords, stopwords("english"))
> him_clean <- tm_map(him_clean, removeNumbers)
> him_clean <- tm_map(him_clean, stripWhitespace)
> him_clean <- tm_map(him_clean, removeWords, c("himalaya", "earthquake"))
```

Bigger words are arranged



Rainbow colors

```
wordcloud(him_clean,random.order = F, scale = c(6,.5))
```

```
wordcloud(him_clean,random.order = F, scale = c(5,2.5))
```

color rainbow effect

```
wordcloud(him_clean,random.order = F, colors = rainbow(50))
```

Rainbow colors

defining max number of words

```
wordcloud(him_clean,random.order = F,
max.words=40, colors = rainbow(50))
```

Rainbow effect