**TWITTER DATA ANALYSIS**

**IMPLEMENTATION**

**STEP 1: Loading packages and registering Oauth.**

require(twitteR)

Loading required package: twitteR

> require(wordcloud)

Loading required package: wordcloud

Loading required package: RColorBrewer

> require(tm)

Loading required package: tm

Loading required package: NLP

> require(stringr)

Loading required package: stringr

> setup\_twitter\_oauth(ConsumerKey, ConsumerSecret, AccessToken, AccessSecret)

[1] "Using direct authentication"

Use a local file to cache OAuth access credentials between R sessions?

1: Yes

2: No

Selection: 1

Error in check\_twitter\_oauth() : OAuth authentication error:

This most likely means that you have incorrectly called setup\_twitter\_oauth()'

> setup\_twitter\_oauth(ConsumerKey, ConsumerSecret, AccessToken, AccessSecret)

[1] "Using direct authentication"

**STEP 2: Extracting tweets (date: from 24-02-2016 to 25-02-2016)**

epl <- searchTwitter("EPL", n=100, lang="en", since = "2016-02-24", until = "2016-02-25")

> mufc <- searchTwitter("Manutd", n=100, lang="en", since = "2016-02-24", until = "2016-02-25")

> cfc <- searchTwitter("Chelsea", n=100, lang="en", since = "2016-02-24", until = "2016-02-25")

**Sample Output:**

"PremierLeagueX: Italy keep quiet on future of Chelsea target Conte https://t.co/qDbKIT2eH7 #Soccer #EPL"

[[2]]

[1] "vofnzambia: RT @Zambia: Multichoice Offers EPL, La Liga On Compact https://t.co/MAWc0qQLK5"

[[3]]

[1] "csgo\_guru: #EPL continues with another Tier-1 match: @astralisgg VS @TeamVirtuspro!\n\nOVERPASS: https://t.co/cYNuCCqwWE\nCACHE: https://t.co/a7J8WuZ90c"

[[4]]

[1] "OsoBipolar7: RT @alshmlaini: @tjpothuraju @TrollFootball probably if Granada was playing in Epl they will be level with Leicester city"

**STEP 3: Creating data frames and extracting only the Hashtags**

epl\_df <- twListToDF(epl)

> mufc\_df <- twListToDF(mufc)

> cfc\_df <- twListToDF(cfc)

>

> epl\_hash <- str\_extract\_all(epl\_df$text, "#\\w+")

> mufc\_hash <- str\_extract\_all(mufc\_df$text, "#\\w+")

> cfc\_hash <- str\_extract\_all(cfc\_df$text, "#\\w+")

> epl\_hash

Sample Output:

[[31]]

[1] "#EPL" "#football" "#twitter92"

[[32]]

character(0)

[[33]]

[1] "#weekender" "#THFC" "#Swans" "#EPL"

[[34]]

character(0)

[[35]]

[1] "#WHUFC" "#SAFC" "#EPL"

**STEP 4: Calculating frequencies of the Hashtags**

epl\_hash <- unlist(epl\_hash)

> mufc\_hash <- unlist(mufc\_hash)

> cfc\_hash <- unlist(cfc\_hash)

>

> epl\_freq <- table(epl\_hash)

> mufc\_freq <- table(mufc\_hash)

> cfc\_freq <- table(cfc\_hash)

> all\_tags <- c(epl\_freq, mufc\_freq, cfc\_freq)

**Sample Output:**

|  |
| --- |
| > cfc\_freq  cfc\_hash  #Atletico #Blues #brunch #cfc #CFC  3 3 1 1 5  #chelsea #Chelsea #football #GaryCahill #Inter  2 3 5 1 1  #Italy #KTBFFH #LFC #nightmare #PFerreira19  1 1 2 1 1  #premierleague #RadioCFC #Saints #Soccer #Southampton  3 1 3 3 3  #spinclasses #theoneshow #WinItWednesday #YR  1 1 4 1 |
|  |
| |  | | --- | | > | |

**STEP 5: Creating wordclouds**

require(wordcloud)

> wordcloud(names(epl\_freq), epl\_freq, random.order=FALSE,

+ colors="#1B9E77")

> title("\n\nHashtags in EPL",

+ cex.main=1.5, col.main="gray50")

Output:

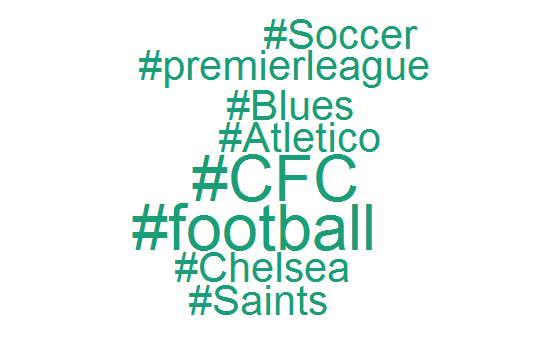


1. wordcloud(names(mufc\_freq), mufc\_freq, random.order=FALSE,
2. + colors="#1B9E77")
3. > title("\n\nHashtags in MUFC",
4. + cex.main=1.5, col.main="gray50")



wordcloud(names(cfc\_freq), cfc\_freq, random.order=FALSE,

+ colors="#1B9E77")



**Aggregate Wordcloud**

# wordcloud

> wordcloud(names(all\_tags), all\_tags, random.order=FALSE, min.freq=1,

+ colors=cols, ordered.colors=TRUE)

> mtext(c("@EPL", "@MUFC", "@CFC"), side=3,

+ line=2, at=c(0.25, 0.5, 0.75), col=c("#1B9E77", "#7570B3", "#D95F02"),

+ family="serif", font=2, cex=1.5)

>

> cols = c(

+ rep("#1B9E77", length(epl\_freq)),

+ rep("#7570B3", length(mufc\_freq)),

+ rep("#D95F02", length(cfc\_freq))

+ )

>

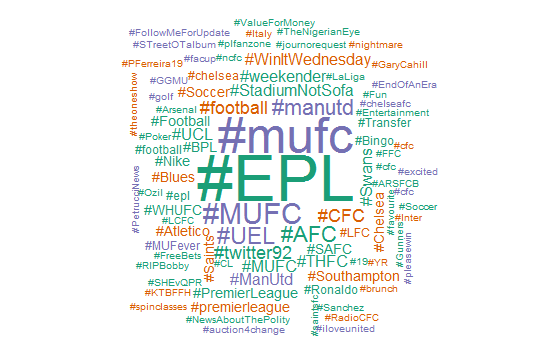
> # wordcloud

> wordcloud(names(all\_tags), all\_tags, random.order=FALSE, min.freq=1,

+ colors=cols, ordered.colors=TRUE)

>

OUTPUT:



**STEP 5: Creating Bar Graphs:**

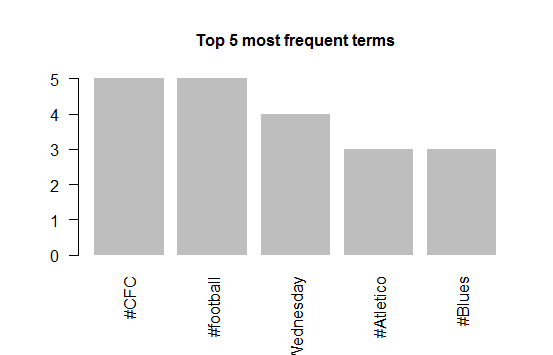
require(ggplot2)

> barplot(table(epl))

top5\_cfc <- head(mfw\_cfc, 5)

> barplot(top5\_cfc, border=NA, las=2, main="Top 5 most frequent terms", cex.main=1)

Output:

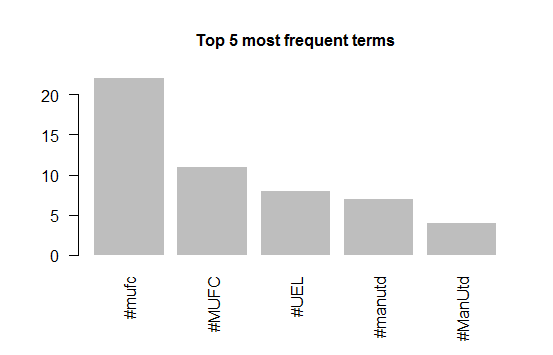


mfw\_mufc <- sort(mufc\_freq, decreasing = TRUE)

> top5\_mufc <- head(mfw\_mufc, 5)

> barplot(top5\_mufc, border=NA, las=2, main="Top 5 most frequent terms", cex.main=1)

Output:



**Overall :**

top <- c(top5\_cfc,top5\_mufc)

> barplot(top, border=NA, las=2, main="Top 5 most frequent terms", cex.main=1)

> barplot(top, border=NA, las=2, main="Most frequent hashtags from 24-2-2016 to 25-02-2016", cex.main=1)

Output:

