Q. 1. Create a wordcloud from the names of the locations of the bike stations, by frequency of usage. Clean up the data by removing words like Avenue, Street, etc. Which names come up often?

⇒ Word cloud for bike stations by frequency of usages.

Fig: 1.0

Data set has been cleaned by removing punctuation, whitespaces, default english words like “the”, “and”, “this” etc…. and following words "lake","&","th","ct","st","avenue","ave","street" are also removed.

On analysing the word cloud based upon the given data set, stations having name “**michigan**”, “**shore**”, “**blvd**” and “**clark**” are the mostly used stations.

Rcode used:

```
install.packages("tm")
install.packages("wordcloud")
install.packages("RColorBrewer")
library(tm)
library(wordcloud)
library(RColorBrewer)

divvyBikeData <-
read.csv("C:/Users/985176/Desktop/WorkingDirectory/Divvy_Stations_Trips_2014_Q1Q2/Divvy
_Trips_2014_Q1Q2_csv.csv")

mycorpus = Corpus(VectorSource(divvyBikeData$from_station_name))

mycorpus = tm_map(mycorpus, removePunctuation)
mycorpus = tm_map(mycorpus, content_transformer(tolower))
mycorpus = tm_map(mycorpus, removeWords,
c("lake","&","th","ct","st","avenue","ave","street",stopwords("english")))
mycorpus = tm_map(mycorpus, stripWhitespace)
mycorpus = tm_map(mycorpus, PlainTextDocument)

tdm <- TermDocumentMatrix(mycorpus)
m1 <- as.matrix(tdm)
v1<- sort(rowSums(m1), decreasing=TRUE)
d1 <- data.frame(word=names(v1), freq=v1)
d1 <- head(d1,100)
wordcloud(d1$word,d1$freq,col=brewer.pal(8,"Set2"), min.freq="5",random.order=FALSE)
```

Q.2. Create a relative importance of the bike stations in the network of stations, using PageRank command. Use every bike rental as a directed link between the From and To stations. Which are the five most important stations?

=> Below fig:2.0 is the directed graph for the mostly used station pair. As shown by the word cloud above in fig 1.0, this shows the top 5 stations that are most frequently used during the trip by the customers or subscribers. The figure also shows the directions of trip in between the stations..

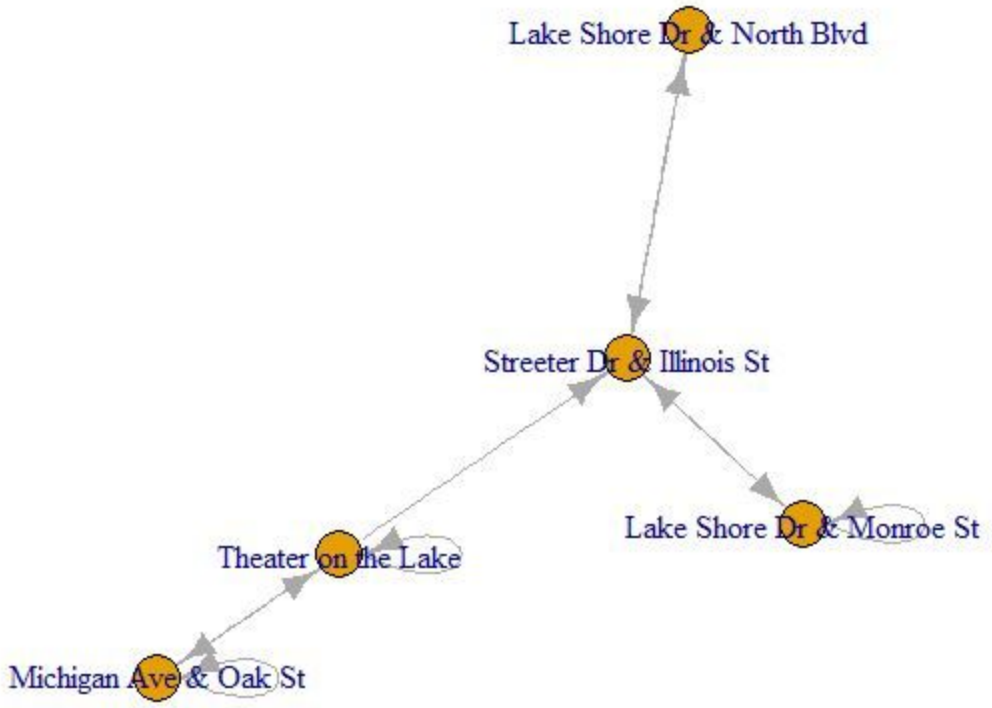

Fig: 2.0

Table show in fig 2.1 is the pagerank calculated for the above stations.

Lake Shore Dr & Monroe St	0.2953864
Michigan Ave & Oak St	0.1028571
Theater on the Lake	0.1028571
Lake Shore Dr & North Blvd	0.1698472
Streeter Dr & Illinois St	0.3290522

Fig: 2.1

Rcode used:

```
library(sqldf)

sortedData <- sqldf("select dv.from_station_name,dv.to_station_name,count(1) from
divvyBikeData dv  group by dv.from_station_id,dv.to_station_id having count(1) > 1 order by
count(1) desc limit 10")

# View(sortedData)
graphData <- graph.data.frame(sortedData,directed = TRUE)

View(page.rank(graphData)$vector)

plot(graphData)
```