Instructions of using code in this project

Note: this file only provides some examples on how to run the code, we also created a html file called project.html to illustrate how these code works in details.

Bubble plot and interactive plot:

To create a bubble plot: import the library ggplot2 and dplyr in R, using the following format of code to accomplish:

```
df_state_avgStar %>% ggplot(aes(x=STATE,y=AVG_STAR,
color=STATE,size=AVG_STAR)) + geom_point() +
theme(axis.text.x=element_text(angle=90, size=20, vjust=0.5)) +
theme(axis.text.x=element_text(size=10, face="bold", vjust=1)) +
theme(axis.title.x=element_text(color="forestgreen",
vjust=0.35),axis.title.y=element_text(color="cadetblue", vjust=0.35)) +
labs(title="State VS BusinessStar",y="BusinessStar",x="State")
```

To create an interactive plot, Import library rCharts, using following syntax:

```
rPlot(AVG_STAR ~ STATE, data = df_state_avgStar, color = 'STATE',
type = 'point')
```

Word Cloud

Using following syntax to create a word cloud:

```
AZ_review_text_tb <- read.csv("~/Desktop/Yelp
Project/Yelp_project_review_dataset/AZ_REVIEWTEXT_DATA_TABLE.csv")
AZ_text <- Corpus(VectorSource(AZ_review_text_tb$TEXT))
AZ_text <- tm_map(AZ_text, stripWhitespace)
AZ_text <- tm_map(AZ_text, tolower)
AZ_text <- tm_map(AZ_text, removeWords, stopwords('english'))
AZ_text <- tm_map(AZ_text, removeWords, c('the'))

wordcloud(AZ_text, scale=c(5,0.5), max.words=50, random.order=FALSE, rot.per=0.1,
use.r.layout=FALSE, colors=brewer.pal(8, 'Dark2'))
```

Dendrogram

Using following syntax to create a dendrogram:

```
### Hierarchical Clustering
```

```
AZ_text_tmDoc <- TermDocumentMatrix(AZ_text,control=list(wordLength=c(1,Inf)))
AZ_text.tdm2 <- removeSparseTerms(AZ_text_tmDoc,sparse = 0.95)
AZ_text.mat2 <- as.matrix(AZ_text.tdm2)
distMat <- dist(scale(AZ_text.mat2))
AZ_text.fit <- hclust(distMat,method='ward')
plot(AZ_text.fit,cex=0.9,hang=-1,main='Word Cluster Dendrogram')
rect.hclust(AZ_text.fit,k=5)
```