Text Mining in R

Contents

[PART 1: Section A 2](#_Toc99729052)

[PART 1: Section B 9](#_Toc99729053)

[Part 2: Section A 13](#_Toc99729054)

[Part 2: Section B 17](#_Toc99729055)

## PART 1: Section A

install.packages(c("mnormt", "psych", "SnowballC", "hunspell",

"broom", "tokenizers", "janeaustenr"))

install.packages(c("slam"))

install.packages("wordcloud")

install.packages("RColorBrewer")

library(ggplot2)

library(dplyr)

library(tidytext)

library(wordcloud)

library(topicmodels)

library(tm)

library(RColorBrewer)

# Reading in the dataset

dataset <- read.csv("Part-1.csv", header=TRUE, encoding="UTF-8")

str(dataset)

#Total tweets distribution as negative,positive,nuetral

ggplot(dataset, aes(x = airline\_sentiment))+

geom\_bar(stat = "count")+

geom\_text(stat='count', aes(label=..count..), vjust=-1)

######## Filtering the dataset into positive, negative & Neutral

positive <- dataset %>%

filter(airline\_sentiment == "positive")

negative <- dataset %>%

filter(airline\_sentiment == "negative")

neutral <- dataset %>%

filter(airline\_sentiment == "neutral")

####### Creating a corpus for each

positive\_corpus <- Corpus(VectorSource(positive$text))

negative\_corpus <- Corpus(VectorSource(negative$text))

neutral\_corpus <- Corpus(VectorSource(neutral$text))

##### Cleaning the text in each corpus

# Cleaning Positive Corpus

toSpace <- content\_transformer(function (x , pattern ) gsub(pattern, " ", x))

positive\_corpus <- tm\_map(positive\_corpus, toSpace, "[^\x01-\x7F]")

positive\_corpus <- tm\_map(positive\_corpus, toSpace, "/")

positive\_corpus <- tm\_map(positive\_corpus, toSpace, "@")

positive\_corpus <- tm\_map(positive\_corpus, toSpace, "\\|")

positive\_corpus <- tm\_map(positive\_corpus, tolower)

positive\_corpus <- tm\_map(positive\_corpus, removePunctuation)

positive\_corpus <- tm\_map(positive\_corpus, removeNumbers)

positive\_corpus <- tm\_map(positive\_corpus, removeWords, stopwords("english"))

positive\_corpus <- tm\_map(positive\_corpus, removeWords, c("will", "http","tco", "flight", "unit","get","can","united"))

positive\_corpus <- tm\_map(positive\_corpus, stemDocument)

positive\_corpus <- tm\_map(positive\_corpus, stripWhitespace)

inspect(positive\_corpus[10:20])

#Cleaning Negative Corpus

toSpace\_neg <- content\_transformer(function (x , pattern ) gsub(pattern, " ", x))

negative\_corpus <- tm\_map(negative\_corpus, toSpace\_neg, "[^\x01-\x7F]")

negative\_corpus <- tm\_map(negative\_corpus, toSpace\_neg, "/")

negative\_corpus <- tm\_map(negative\_corpus, toSpace\_neg, "@")

negative\_corpus <- tm\_map(negative\_corpus, toSpace\_neg, "\\|")

negative\_corpus <- tm\_map(negative\_corpus, tolower)

negative\_corpus <- tm\_map(negative\_corpus, removePunctuation)

negative\_corpus <- tm\_map(negative\_corpus, removeNumbers)

negative\_corpus <- tm\_map(negative\_corpus, removeWords, stopwords("english"))

negative\_corpus <- tm\_map(negative\_corpus, removeWords, c("will", "http","tco", "flight", "unit","get","can","united"))

negative\_corpus <- tm\_map(negative\_corpus, stemDocument)

negative\_corpus <- tm\_map(negative\_corpus, stripWhitespace)

inspect(negative\_corpus[1:10])

#Cleaning Neutral Corpus

toSpace\_ne <- content\_transformer(function (x , pattern ) gsub(pattern, " ", x))

neutral\_corpus <- tm\_map(neutral\_corpus, toSpace\_ne, "[^\x01-\x7F]")

neutral\_corpus <- tm\_map(neutral\_corpus, toSpace\_ne, "/")

neutral\_corpus <- tm\_map(neutral\_corpus, toSpace\_ne, "@")

neutral\_corpus <- tm\_map(neutral\_corpus, toSpace\_ne, "\\|")

neutral\_corpus <- tm\_map(neutral\_corpus, tolower)

neutral\_corpus <- tm\_map(neutral\_corpus, removePunctuation)

neutral\_corpus <- tm\_map(neutral\_corpus, removeNumbers)

neutral\_corpus <- tm\_map(neutral\_corpus, removeWords, stopwords("english"))

neutral\_corpus <- tm\_map(neutral\_corpus, removeWords, c("will", "http","tco", "flight", "unit","get","can","united"))

neutral\_corpus <- tm\_map(neutral\_corpus, stemDocument)

neutral\_corpus <- tm\_map(neutral\_corpus, stripWhitespace)

inspect(neutral\_corpus[1:10])

# Pre-processing ends

# Document Term matrix: Represention

dtm <- DocumentTermMatrix(positive\_corpus)

dtm\_neg <- DocumentTermMatrix(negative\_corpus)

dtm\_ne <- DocumentTermMatrix(neutral\_corpus)

# Removing empty rows

rowTotals <- apply(dtm , 1, sum) #Find the sum of words in each Document

dtm.new <- dtm[rowTotals> 0, ]

rowTotals\_neg <- apply(dtm\_neg , 1, sum) #Find the sum of words in each Document

dtm.new\_neg <- dtm\_neg[rowTotals\_neg> 0, ]

rowTotals\_ne <- apply(dtm\_ne , 1, sum) #Find the sum of words in each Document

dtm.new\_ne <- dtm\_ne[rowTotals\_ne> 0, ]

##############################################

#Applying LDA, displaying results

########### LDA Positive

# LDA Positive Two Topics

pos\_lda <- LDA(dtm.new, k = 2, control = list(seed = 1234))

terms(pos\_lda, 20)

pos\_topics <- tidy(pos\_lda, matrix = "beta")

pos\_topics

pos\_top\_terms <- pos\_topics %>%

group\_by(topic) %>%

slice\_max(beta, n = 10) %>%

ungroup() %>%

arrange(topic, -beta)

pos\_top\_terms %>%

mutate(term = reorder\_within(term, beta, topic)) %>%

ggplot(aes(beta, term, fill = factor(topic))) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~ topic, scales = "free") +

scale\_y\_reordered() +

labs(title = "Positive Latent Dirichlet Allocation with Two Topics") +

labs(y = "Word") +

labs(x = "Beta")

# LDA Positive 5 Topics

pos\_lda <- LDA(dtm.new, k = 5, control = list(seed = 1234))

terms(pos\_lda, 20)

pos\_topics <- tidy(pos\_lda, matrix = "beta")

pos\_topics

pos\_top\_terms <- pos\_topics %>%

group\_by(topic) %>%

slice\_max(beta, n = 10) %>%

ungroup() %>%

arrange(topic, -beta)

pos\_top\_terms %>%

mutate(term = reorder\_within(term, beta, topic)) %>%

ggplot(aes(beta, term, fill = factor(topic))) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~ topic, scales = "free") +

scale\_y\_reordered() +

labs(title = "Positive Latent Dirichlet Allocation with Five Topics") +

labs(y = "Word") +

labs(x = "Beta")

#################### LDA Negative

neg\_lda <- LDA(dtm.new\_neg, k = 2, control = list(seed = 1234))

terms(neg\_lda, 20)

neg\_topics <- tidy(neg\_lda, matrix = "beta")

neg\_topics

neg\_top\_terms <- neg\_topics %>%

group\_by(topic) %>%

slice\_max(beta, n = 10) %>%

ungroup() %>%

arrange(topic, -beta)

neg\_top\_terms %>%

mutate(term = reorder\_within(term, beta, topic)) %>%

ggplot(aes(beta, term, fill = factor(topic))) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~ topic, scales = "free") +

scale\_y\_reordered() +

labs(title = "Negative Latent Dirichlet Allocation with Two Topics") +

labs(y = "Word") +

labs(x = "Beta")

# LDA Negative 5 Topics

neg\_lda <- LDA(dtm.new\_neg, k = 5, control = list(seed = 1234))

terms(neg\_lda, 20)

neg\_topics <- tidy(neg\_lda, matrix = "beta")

neg\_topics

neg\_top\_terms <- neg\_topics %>%

group\_by(topic) %>%

slice\_max(beta, n = 10) %>%

ungroup() %>%

arrange(topic, -beta)

neg\_top\_terms %>%

mutate(term = reorder\_within(term, beta, topic)) %>%

ggplot(aes(beta, term, fill = factor(topic))) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~ topic, scales = "free") +

scale\_y\_reordered() +

labs(title = "Negative Latent Dirichlet Allocation with Five Topics") +

labs(y = "Word") +

labs(x = "Beta")

############################# LDA Neutral

# LDA Neutral Two Topics

ne\_lda <- LDA(dtm.new\_ne, k = 2, control = list(seed = 1234))

terms(ne\_lda, 20)

ne\_topics <- tidy(ne\_lda, matrix = "beta")

ne\_topics

ne\_top\_terms <- ne\_topics %>%

group\_by(topic) %>%

slice\_max(beta, n = 10) %>%

ungroup() %>%

arrange(topic, -beta)

ne\_top\_terms %>%

mutate(term = reorder\_within(term, beta, topic)) %>%

ggplot(aes(beta, term, fill = factor(topic))) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~ topic, scales = "free") +

scale\_y\_reordered() +

labs(title = "Neutral Latent Dirichlet allocation with Two Topics") +

labs(y = "Word") +

labs(x = "Beta")

# LDA Neutral 5 Topics

ne\_lda <- LDA(dtm.new\_ne, k = 5, control = list(seed = 1234))

terms(ne\_lda, 20)

ne\_topics <- tidy(ne\_lda, matrix = "beta")

ne\_topics

ne\_top\_terms <- ne\_topics %>%

group\_by(topic) %>%

slice\_max(beta, n = 10) %>%

ungroup() %>%

arrange(topic, -beta)

ne\_top\_terms %>%

mutate(term = reorder\_within(term, beta, topic)) %>%

ggplot(aes(beta, term, fill = factor(topic))) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~ topic, scales = "free") +

scale\_y\_reordered() +

labs(title = "Neutral Latent Dirichlet Allocation with Five Topics") +

labs(y = "Word") +

labs(x = "Beta")

########################### Part B #######################

#Part 1.B: Script

full\_dataset <- Corpus(VectorSource(dataset$text))

inspect(full\_dataset)

#cleaning full dataset

toSpace\_full <- content\_transformer(function (x , pattern ) gsub(pattern, " ", x))

full\_dataset <- tm\_map(full\_dataset, toSpace\_full, "[^\x01-\x7F]")

full\_dataset <- tm\_map(full\_dataset, toSpace, "/")

full\_dataset <- tm\_map(full\_dataset, toSpace, "@")

full\_dataset <- tm\_map(full\_dataset, toSpace, "\\|")

full\_dataset <- tm\_map(full\_dataset, tolower)

full\_dataset <- tm\_map(full\_dataset, removePunctuation)

full\_dataset <- tm\_map(full\_dataset, removeNumbers)

full\_dataset <- tm\_map(full\_dataset, removeWords, stopwords("english"))

full\_dataset <- tm\_map(full\_dataset, removeWords, c("will", "http","tco", "flight", "unit","get","can","united"))

full\_dataset <- tm\_map(full\_dataset, stemDocument)

full\_dataset <- tm\_map(full\_dataset, stripWhitespace)

inspect(full\_dataset[1:10])

# Document Term matrix: Represention

dtm\_full <- DocumentTermMatrix(full\_dataset)

#

memory.limit(size = 56000)

#

# Removing empty rows

rowTotals\_full <- apply(dtm\_full , 1, sum) #Find the sum of words in each Document

dtm.new\_full <- dtm\_full[rowTotals\_full> 0, ]

#visualise

full\_lda <- LDA(dtm.new\_full, k = 3, control = list(seed = 1234))

terms(full\_lda, 20)

full\_topics <- tidy(full\_lda, matrix = "beta")

full\_topics

full\_top\_terms <- full\_topics %>%

group\_by(topic) %>%

slice\_max(beta, n = 10) %>%

ungroup() %>%

arrange(topic, -beta)

full\_top\_terms %>%

mutate(term = reorder\_within(term, beta, topic)) %>%

ggplot(aes(beta, term, fill = factor(topic))) +

geom\_col(show.legend = FALSE) +

facet\_wrap(~ topic, scales = "free") +

scale\_y\_reordered() +

labs(title = "Latent Dirichlet Allocation with three Topics (Full Dataset)") +

labs(y = "Word") +

labs(x = "Beta")

## PART 1: Section B

* From analysis of the tweets, it was identified that 9,178 were negative, 3,099 were neutral and 2,363 were positive.

Chart, bar chart, funnel chart

Description automatically generated

* Topic one in the graph above may refer to the good service provided by JetBlue in particular.
* Topic two refers to good experiences with the airlines southwest Air and US Airways with ‘great’ and ‘awesome’ being frequently used words.

Chart, waterfall chart

Description automatically generated with medium confidence

* In general, the main theme seen among all topics seems to be tweets thanking airlines for the good experience they provided.
* Some of the frequently used words from the tweets in the positive sentiment included ' thank',’jetblu’, ‘southwestair’, 'great', 'love', 'best' and 'awesome'.

Chart, funnel chart

Description automatically generated

* Topic one may refer to the negative experiences with Jetblue in particuliar because it is the most frequent word used.
* Topic two on the other hand may refer to the negative experiences had with American Air in pariculiar.
* Both seeem to refer to delays.

Chart, bar chart

Description automatically generated

* Some of the frequently used words from the tweets in the negative sentiment included ’jetblu’, ‘southwestair’,’ameriacanair’, ‘cancel', 'wait', 'time' and 'usairway'.
* Topic one: Centres around delays.
* Topic two: Negative sentiment towards US Airways. This may be due to the loss of ‘bag’ or customers unable to ‘make’ their ‘flights’ due to holdups or customs.
* These topics are mainly centered around ‘delays’ and cancellations.

Chart, funnel chart

Description automatically generated

* Topic one from the two topic LDA on neutral sentiment seems to be about customers needing stuff from the airlines. This may be a change or cancel their flights.
* Topic 2 seems to refer to help needed in regards to flights.

Chart

Description automatically generated

* Some of the frequently used words from the tweets in the neutral sentiment sentiment included ’jetblu’, ‘southwestair’,’ameriacanair’, ‘tommorrow', 'need', 'please' and 'usairway'.

Chart, bar chart, funnel chart

Description automatically generated

* Some of the frequently used words from the tweets from the whole dataset included ’jetblu’, ‘southwestair’,’ameriacanair’, ‘hour', 'delay', 'cancel', ‘nned’ and 'usairway'.
* From performing an LDA on the dataset as a whole the three sentiments, positive, negative and neutral can be clearly identified from the three topics in the graph above.
* Topic 1 is quite similar in terms of words to the Negative LDA.
* Topic 3 is quite similar in terms of words to the Neutral LDA.

## Part 2: Section A

library(ggplot2)

library(dplyr)

library(tidytext)

library(wordcloud)

library(topicmodels)

library(tm)

library(RColorBrewer)

library(dplyr, quietly = T)

Q2Dataset <- read.csv("Part-2.csv", header=TRUE, encoding="UTF-8")

# Exploring the dataset

head(Q2Dataset)

summary(Q2Dataset)

# Cleaning the dataset

rem\_reg <- "&amp;|&lt;|&gt"

Q2\_tidy <- Q2Dataset %>%

mutate(text = str\_remove\_all(text, rem\_reg),

text = gsub("\\s?(f|ht)(tp)(s?)(://)([^\\.]\*)[\\.|/](\\S\*)", "", text),

text = gsub("[^0-9A-Za-z///' ]", "", text),

text = gsub("[[:punct:][:digit:]]", "", text)) %>%

unnest\_tokens(word, text, token = "tweets") %>%

#Change to tokens

anti\_join(stop\_words) %>%

#Eliminate stop words

filter((!word=='rt')|(!word=='amp'))

# creating corpus

Cor <- Corpus(VectorSource(Q2Dataset$text))

Cor <- tm\_map(Cor, content\_transformer(tolower))

Cor <- tm\_map(Cor, content\_transformer(gsub), pattern= "\\W", replace= " ")

rem\_amp <- function(x) gsub("amp","", x)

Cor <- tm\_map(Cor, content\_transformer(rem\_amp))

rem\_non <- function(x) gsub("[^[:alpha:][:space:]]\*","", x)

Cor <- tm\_map(Cor, content\_transformer(rem\_non))

rem\_url <- function(x) gsub("http[^[:space:]]\*","", x)

Cor <- tm\_map(Cor, content\_transformer(rem\_url))

rem\_emojis <- function(x) gsub("[^0-9A-Za-z///' ]","", x)

Cor <- tm\_map(Cor, content\_transformer(rem\_emojis))

Cor <- tm\_map(Cor, removeWords, stopwords("english"))

Cor <- tm\_map(Cor, removePunctuation)

Cor <- tm\_map(Cor, removeNumbers)

Cor <- tm\_map(Cor, stripWhitespace)

# Creating a document term matrix

dtm <- TermDocumentMatrix(Cor)

dtm\_matrix <- as.matrix(dtm)

sorting\_row <- sort(rowSums(dtm\_matrix), decreasing = TRUE)

df <- data.frame(word = names(sorting\_row), n=sorting\_row)

##############################################################

# a) A frequency plot of the top 25 words

df %>%

top\_n(25) %>%

mutate(word = reorder(word, n)) %>%

ggplot(aes(x = word, y = n, fill = word)) +

geom\_col(show.legend = FALSE) +

xlab(NULL) +

coord\_flip() +

labs(y = "Count") +

labs(x = "Unique words") +

labs(title = "The Top 25 Most Frequent Words") +

theme(legend.position = "none") +

theme\_bw()

##################################################################

# b) A word cloud of 40 most common words

set.seed(1234)

tweet\_40\_most\_common = tweet\_stop\_words %>%

count(word, sort = TRUE) %>%

top\_n(40) %>%

mutate(word = reorder(word, n))

wordcloud(words=tweet\_40\_most\_common$word, freq=tweet\_40\_most\_common$n, min.freq=1, scale=c(2.5, .5),

max.words=50, random.order=FALSE, rot.per=0.35,

colors=brewer.pal(8, "Dark2"))

##################################################################

#c) Can you check he was using Social media during breakfast time (6-10 AM)

# or not? (Hint: Plot by the hour)

Q2Dataset$Date <- as.Date(Q2Dataset$date)

Q2Dataset$Time <- format(as.POSIXct(Q2Dataset$date), format="%H:%M")

tweets\_by\_time = Q2Dataset %>%

filter(Time >= "06:00" & Time <= "10:00")

tweets\_by\_date = tweets\_by\_time %>%

group\_by(tweets\_by\_time$Time) %>%

count()

sixtoten <- tweets\_by\_time$Time

hours <- as.numeric( format( strptime( sixtoten , format = "%H:%M" ) , "%H" ) )

hist(hours ,

breaks = unique( hours ),

xlab = 'Hours',

main = "Number of Tweets per hour between 6am-10am",

col = "dodgerblue3")

########################################################

# d) What was the usage per month?

Q2Dataset$Month <- format(Q2Dataset$Date, "%Y-%m")

tweets\_months = Q2Dataset %>%

group\_by(Month) %>%

count()

ggplot(tweets\_months, aes(x = Month, y = n, group = 1)) +

geom\_line() +

theme(axis.text.x = element\_text(angle=90)) +

labs(y = "Number of Tweets") +

labs(title = "Number of Tweets per Month (Jan 2017 - Sept 2018)")

#######################################################

# e) What were the top-15 words for source=’iPhone’ and source=’Media Studio’.

# Any interesting about this?

# Do think both sources were handled by one person only?

data <- read.csv("Part-2.csv", header=TRUE, encoding="UTF-8")

data$text <- tolower(data$text)

data$text <- gsub("rt", "", data$text) # RT

data$text <- gsub("https\\S\*", "", data$text) #links

data$text <- gsub("[[:punct:]]", "", data$text) #punctuation

data$text <- gsub("@\\S\*", "", data$text) #mentions

data$text <- gsub("[\r\n]", "", data$text) #\r\n\

data$text <- gsub("amp", "", data$text) # amp

# Deleting stop-words

tweet\_stop\_words <- data %>%

select(text) %>%

unnest\_tokens(word, text)

tweet\_stop\_words <- tweet\_stop\_words %>%

anti\_join(stop\_words)

tweets\_iphone = data %>%

filter(source == "Twitter for iPhone")

tweets\_media = data %>%

filter(source == "Media Studio")

tweet\_words\_iphone <- tweets\_iphone %>%

select(text) %>%

unnest\_tokens(word, text)

tweet\_words\_iphone <- tweet\_words\_iphone %>%

anti\_join(stop\_words)

tweet\_words\_iphone %>%

count(word, sort = TRUE) %>%

top\_n(15)

tweet\_words\_media <- tweets\_media %>%

select(text) %>%

unnest\_tokens(word, text)

tweet\_words\_media <- tweet\_words\_media %>%

anti\_join(stop\_words)

tweet\_words\_media %>%

count(word, sort = TRUE) %>%

top\_n(15)

#######################################

# f) What are the six words that he did not use in the last six months of the data

# but were frequently used in the first six months?

first\_six <- data %>%

filter(Date >= "2017-01-20" & Date <= "2017-07-20")

last\_six <- data %>%

filter(Date >= "2018-03-01" & Date <= "2018-09-01")

tweet\_first\_6 <- first\_six %>%

select(text) %>%

unnest\_tokens(word, text)

tweet\_first\_6 <- tweet\_first\_6 %>%

anti\_join(stop\_words)

tweet\_first\_6 %>%

count(word, sort = TRUE) %>%

top\_n(15)

tweet\_last\_6 <- last\_six %>%

select(text) %>%

unnest\_tokens(word, text)

tweet\_last\_6 <- tweet\_last\_6 %>%

anti\_join(stop\_words)

tweet\_last\_6 %>%

count(word, sort = TRUE) %>%

top\_n(15)

## Part 2: Section B

Chart, funnel chart

Description automatically generated

* It can be seen from the graph above that ‘great’ is the most common word used in Mr. XYZ’s tweets.
* This word is used over 1000 times
* The most common word used ‘great’ is tweeted over twice as much as the second most common word ‘People’.
* Many words relating to American politics are tweeted including Trump, president, democrats, tax, and jobs.
* This may imply that Mr. XYZ is an American citizen or that he has a career in politics or involved in politics in some other form.

Bar chart

Description automatically generated with medium confidence

Chart, histogram

Description automatically generated

* The histogram above analyses Mr. XYZ’s tweeting activity in the morning.
* Mr. XYZ is very active early in the morning with over twenty-five tweets between six and seven.
* Majority of his tweeting activity occurs between 8am and 9am.
* After 9pm we can see a dip in activity perhaps this is just before he goes to work where he may not be on his devices as much.

Chart, line chart

Description automatically generated

* The graph above analyses the number of tweets made per month from January 2017 until September 2018.
* Over this period of twenty months the greatest number of tweets made was in July 2018.
* In July 2018 close to four hundred tweets were made by Mr. XYZ.
* A noticeable trend from the graph is that Mr. XYZ tweeting activity tends to drop in the first few months of the year.
* Around one-hundred and fifty tweets are made in January, February and March for both 2017 & 2018.
* After these months pass an upwards trend in activity usually ensues.

Top Fifteen words for iPhone Top 15 words for Media Studio

Text

Description automatically generated Text

Description automatically generated

* Both sources contain quite similar topics.
* Mainly both sources centre around America and politics.
* Tweets made in the studio have a more patriotic tone.
* However, the words tweeted in the studio are far less in size compared to those made on the phone.
* This may be due to a mobile device being much more convenient.
* The media studio perhaps could be a place of work for Mr. XYZ therefore he would not have time to tweet as much.

**Frequently used first six months** **Frequently used last six months**

A screenshot of a computer

Description automatically generated with low confidence Text

Description automatically generated

* From analysis of the of the most frequently used words from the first six months and the last six months popular words like ‘jobs’, ‘America’, ‘American’, ‘fox and friends’, ‘election’ and ‘Obamacare’ were not used in the last six months.
* This may signify the political topics that were prominent in the earlier tweets produced that were not as relevant the following year.