Data Visualization Lab Assignment

Name: Om Ashish Mishra

Registration Number: 16BCE0789

Slot: C2

Q) Take a text document. Show the word cloud for the document.

```
> library("tm")
> library("SnowballC")
> library("wordcloud")
> library("RColorBrewer")
> text <- readLines(file.choose())
Warning message:
In readLines(file.choose()) :
  incomplete final line found on 'C:\Users\Aditya Dhall\Desktop\text.txt'
> docs <- Corpus(VectorSource(text))
> inspect(docs)
<<SimpleCorpus>>
Metadata:  corpus specific: 1, document level (indexed): 0
Content:  documents: 3

[1] Lorem ipsum is a pseudo-Latin text used in web design, typography,
layout, and printing in place of English to emphasise design elements over
content. It's also called placeholder (or filler) text. It's a convenient
tool for mock-ups. It helps to outline the visual elements of a document
or presentation, eg typography, font, or layout. Lorem ipsum is mostly a
part of a Latin text by the classical author and philosopher Cicero. Its
words and letters have been changed by addition or removal, so to
deliberately render its content nonsensical; it's not genuine, correct, or
comprehensible Latin anymore. While lorem ipsum's still resembles
classical Latin, it actually has no meaning whatsoever. As Cicero's text
doesn't contain the letters K, W, or Z, alien to latin, these, and others
are often inserted randomly to mimic the typographic appearence of
European languages, as are digraphs not to be found in the original.
[2]
[3] In a professional context it often happens that private or corporate
clients corder a publication to be made and presented with the actual
content still not being ready. Think of a news blog that's filled with
content hourly on the day of going live. However, reviewers tend to be
distracted by comprehensible content, say, a random text copied from a
newspaper or the internet. The are likely to focus on the text,
disregarding the layout and its elements. Besides, random text risks to be
unintendedly humorous or offensive, an unacceptable risk in corporate
environments. Lorem ipsum and its many variants have been employed since
the early 1960ies, and quite likely since the sixteenth century.
> toSpace <- content_transformer(function (x , pattern ) gsub(pattern, "
", x))
> docs <- tm_map(docs, toSpace, "/")
Warning message:
```

```
In tm_map.SimpleCorpus(docs, toSpace, "/") : transformation drops
documents
> docs <- tm_map(docs, toSpace, "@")
Warning message:
In tm_map.SimpleCorpus(docs, toSpace, "@") : transformation drops
documents
> docs <- tm_map(docs, toSpace, "\\|")
Warning message:
In tm_map.SimpleCorpus(docs, toSpace, "\\|") :
  transformation drops documents
> docs <- tm_map(docs, content_transformer(tolower))
Warning message:
In tm_map.SimpleCorpus(docs, content_transformer(tolower)) :
  transformation drops documents
> docs <- tm_map(docs, removeNumbers)
Warning message:
In tm_map.SimpleCorpus(docs, removeNumbers) :
  transformation drops documents
> docs <- tm_map(docs, removeWords, stopwords("english"))
Warning message:
In tm_map.SimpleCorpus(docs, removeWords, stopwords("english")) :
  transformation drops documents
> docs <- tm_map(docs, removePunctuation)
Warning message:
In tm_map.SimpleCorpus(docs, removePunctuation) :
  transformation drops documents
> docs <- tm_map(docs, stripWhitespace)
Warning message:
In tm_map.SimpleCorpus(docs, stripWhitespace) :
  transformation drops documents
> dtm <- TermDocumentMatrix(docs)
> m <- as.matrix(dtm)
> v <- sort(rowSums(m),decreasing=TRUE)
> d <- data.frame(word = names(v),freq=v)
> head(d, 10)
                          word freq
text                      text    7
content                content    5
latin                    latin    4
lorem                    lorem    4
elements              elements    3
ipsum                    ipsum    3
layout                  layout    3
classical            classical    2
comprehensible comprehensible    2
design                  design    2
>
>
> wordcloud(words = d$word, freq = d$freq, min.freq = 1,max.words=200,
random.order=FALSE, rot.per=0.35,colors=brewer.pal(8, "Dark2"))
```
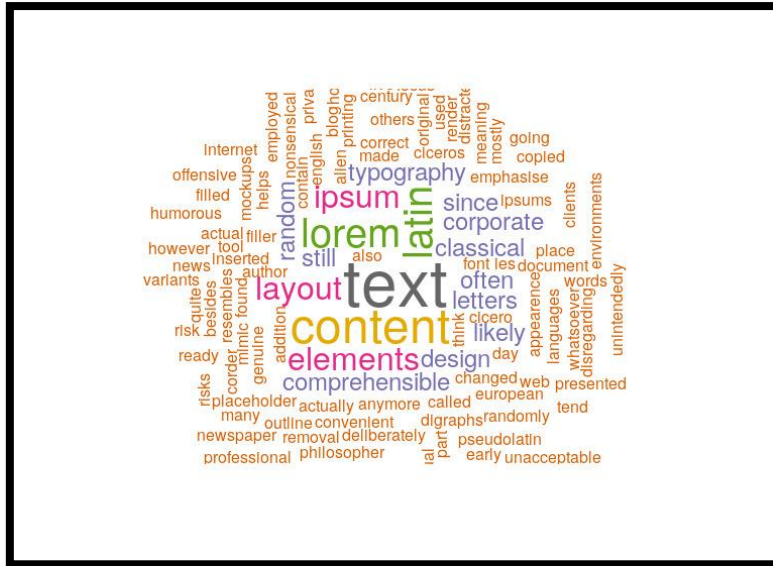
Q2) Take a time series data. Show the three components of the time series data.

```
> library(Quandl)
> Quandl.search("new home construction")
```
Uruguay - New Housing Construction Index
Code: SGE/URYHOUS
Desc: <p>This index tracks the number of new houses built in Uruguay
.</p><p>Units: Index Points 1990=100, NSA</p><p>Source: <a
href =http://www.ine.gub.uy/>Instituto Nacional de Estadstica,
Uruguay</p>
Freq: quarterly
Cols: Date | Value

Housing Market Indicators in New Brunswick: Construction
Code: CMHC/NB_HOUSINGCONSTRUCTION
Desc: These tables bring together information from a variety of sources.
They provide an overview of housing conditions and trends in Canada and in
each province and Census Metropolitan Area. Some tables assemble data for
different markets, while others feature one particular market. Contents
include information on housing construction, and sales; rental market
trends; mortgage lending; the housing stock; household characteristics;
and housing need. Tables also contain data on factors that influence
housing markets, such as employment trends, demographic changes, and
income growth.
Freq: annual
Cols: Year | Total Starts | Single Starts | Multiple Starts | Semi-
Detached | Row | Apartment | Starts Intended Mkt | Owned | Rental | Condo
| Other Starts | Completions

New Privately-Owned Housing Units Under Construction: Total
Code: FRED/UNDCONTNSA
Desc: Thousands of Units Not Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Under Construction: Total

```
Code: FRED/UNDCONTSA
Desc: Thousands of Units Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Under Construction in the West Census
Region
Code: FRED/UNDCONWTSA
Desc: Thousands of Units Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Under Construction in the Midwest Census
Region
Code: FRED/UNDCONMWTSA
Desc: Thousands of Units Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Under Construction in the Northeast
Census Region
Code: FRED/UNDCONNETNSA
Desc: Thousands of Units Not Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Under Construction: 1-Unit Structures
Code: FRED/UNDCON1UNSA
Desc: Thousands of Units Not Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Under Construction in the Midwest Census
Region
Code: FRED/UNDCONMWTNSA
Desc: Thousands of Units Not Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Under Construction in the West Census
Region
Code: FRED/UNDCONWTNSA
Desc: Thousands of Units Not Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

> Quandl.search("Housing Units Completed", source="FRED")
New Privately-Owned Housing Units Completed: Total
Code: FRED/COMPUTSA
Desc: Thousands of Units Seasonally Adjusted Annual Rate,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Completed: Total
Code: FRED/COMPUTNSA
Desc: Thousands of Units Not Seasonally Adjusted,
Freq: monthly
Cols: Date | Value

New Privately-Owned Housing Units Completed: 1-Unit Structures
Code: FRED/COMPU1USA
Desc: Thousands of Units Seasonally Adjusted Annual Rate,
```

```
        Freq: monthly
        Cols: Date | Value

        New Privately-Owned Housing Units Completed: 1-Unit Structures
        Code: FRED/COMPU1UNSA
        Desc: Thousands of Units Not Seasonally Adjusted,
        Freq: monthly
        Cols: Date | Value

        New Privately-Owned Housing Units Completed: 2-4 Unit Structures
        Code: FRED/COMPU24UNSA
        Desc: Thousands of Units Not Seasonally Adjusted,
        Freq: monthly
        Cols: Date | Value

        New Privately-Owned Housing Units Completed: 2-4 Unit Structures
        Code: FRED/COMPU24USA
        Desc: Thousands of Units Seasonally Adjusted Annual Rate,
        Freq: monthly
        Cols: Date | Value

        New Privately-Owned Housing Units Completed: 5-Unit Structures or More
        Code: FRED/COMPU5MUNSA
        Desc: Thousands of Units Not Seasonally Adjusted,
        Freq: monthly
        Cols: Date | Value

        New Privately-Owned Housing Units Completed: 5-Unit Structures or More
        Code: FRED/COMPU5MUSA
        Desc: Thousands of Units Seasonally Adjusted Annual Rate,
        Freq: monthly
        Cols: Date | Value

        New Privately-Owned Housing Units Completed in the Northeast Census Region
        Code: FRED/COMPUNETNSA
        Desc: Thousands of Units Not Seasonally Adjusted,
        Freq: monthly
        Cols: Date | Value

        New Privately-Owned Housing Units Completed in the Midwest Census Region
        Code: FRED/COMPUMWTNSA
        Desc: Thousands of Units Not Seasonally Adjusted,
        Freq: monthly
        Cols: Date | Value
> Units = Quandl("FRED/COMPUTSA")
> head(Units)
        Date Value
1 2019-02-01  1303
2 2019-01-01  1247
3 2018-12-01  1063
4 2018-11-01  1101
5 2018-10-01  1111
6 2018-09-01  1148
> value.ts = ts(Value, data=Units, frequency=12, start=c(2000,1),
end=c(2018,12))
> plot(value.ts)
```
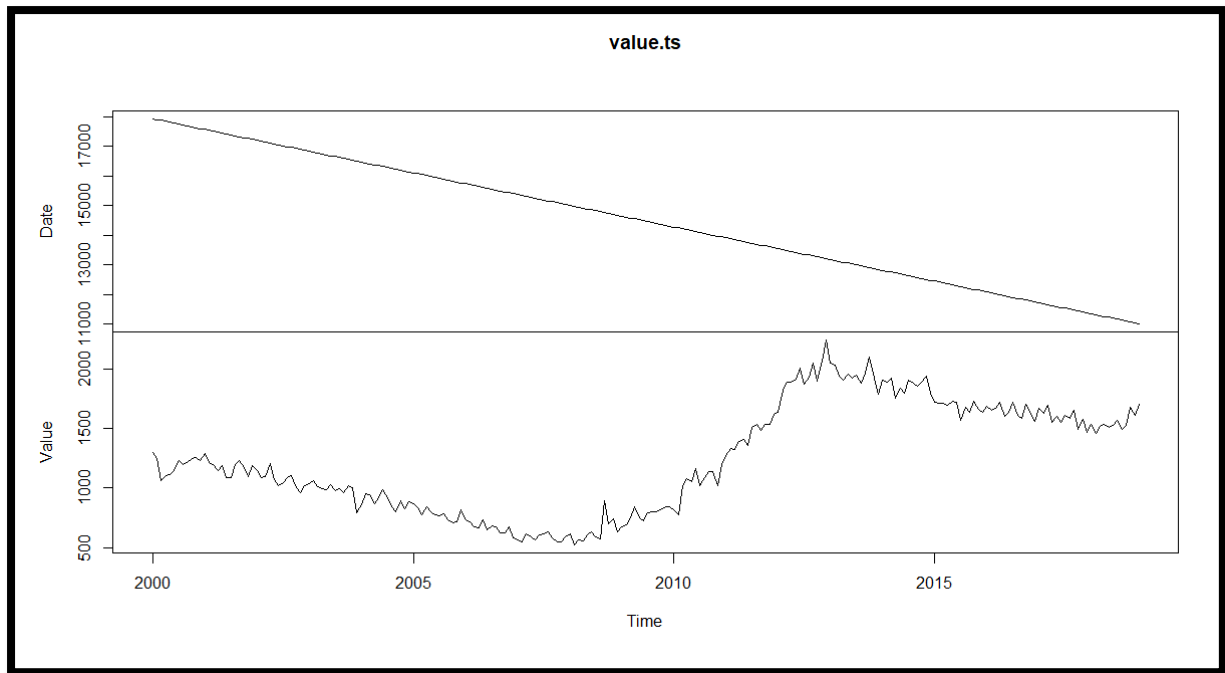
value.ts

```
> d = decompose(value.ts)
> plot(d)
```


Decomposition of additive time series