Libraries

```
library(ggplot2)          # ggplot()
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
library(mongolite)
```

```
## Warning: package 'mongolite' was built under R version 4.1.3
```

```
library(GGally)           # ggcorr() | ggpairs()
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
## For corpus creation
library(stringr)
library(bitops)
library(NLP)              # as.String()
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
library(tm)

## for lemantizing wordcloud and graph words
library(pacman)
pacman::p_load_gh("trinker/textstem")
pacman::p_load(textstem, dplyr)

## For Word Cloud
library(RColorBrewer)
```

```
## Warning: package 'RColorBrewer' was built under R version 4.1.3
```

```
library(wordcloud)

## For clustering of words
#library(graph)

## For sentiments  analysis
library(syuzhet)
```

```
## Warning: package 'syuzhet' was built under R version 4.1.3
```

```
library(scales)            #blank_theme()
```

```
## Warning: package 'scales' was built under R version 4.1.3
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:syuzhet':
##
##     rescale
```

```
## Color Pallette
pallette = brewer.pal(8,"Dark2")
```

Connect to MongoDB

```
connection = 'mongodb://localhost:27017'
hotelreviews = mongo(collection = "reviews",
                     db = "HotelReviews",
                     url = connection)
hotelreviews$count()
```

```
## [1] 2230
```

Data Prep

```
df = hotelreviews$find(sort = '{"positive" : -1}',
                       fields = '{"Name" : true,
                                  "Contribution" : true,
                                  "Rating" : true,
                                  "Review" : true,
                                  "TripType" : true,
                                  "HotelName" : true,
                                  "City" : true}')
str(df)
```

```
## 'data.frame':    2230 obs. of  7 variables:
##  $ _id      : chr  "62db783847fdb0fdd2afd7db" "62db783847fdb0fdd2afd7dc" "62db783847fdb0fdd2afd7dd" "62db78384
7fdb0fdd2afd7de" ...
##  $ Name     : chr  "Summernole96" "beatrix659" "williamsS293YZ" "LizMSK" ...
##  $ Rating   : chr  "45" "45" "50" "45" ...
##  $ Review   : chr  "Great hotel for your trip to Paris Such a wonderful place to stay on our first visit to Pa
ris as a family. Ever"| __truncated__ "Lovely staying in Paris We stayed in this gorgeous little hotel with my fi
ance for a weekend in Paris. It has a"| __truncated__ "Charming, great location, great service Clean, charming ho
tel located near the wonderful Rue Cler, a typical Fr"| __truncated__ "Intimate hotel with much to offer! I staye
d for 8 nights in late May 2022 at the Hotel de la Motte Picquet.  My"| __truncated__ ...
##  $ TripType : chr  "Trip type: Travelled with family" "Trip type: Travelled as a couple" "Trip type: Travelled
with friends" "Trip type: Travelled with family" ...
##  $ HotelName: chr  "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" ...
##  $ City     : chr  "There are more places to choose from in the Paris area." "There are more places to choose
from in the Paris area." "There are more places to choose from in the Paris area." "There are more places to choo
se from in the Paris area." ...
```

```
# Omit ID column
df = df[-c(1)]
str(df)
```

```
## 'data.frame':    2230 obs. of  6 variables:
##  $ Name     : chr  "Summernole96" "beatrix659" "williamsS293YZ" "LizMSK" ...
##  $ Rating   : chr  "45" "45" "50" "45" ...
##  $ Review   : chr  "Great hotel for your trip to Paris Such a wonderful place to stay on our first visit to Pa
ris as a family. Ever"| __truncated__ "Lovely staying in Paris We stayed in this gorgeous little hotel with my fi
ance for a weekend in Paris. It has a"| __truncated__ "Charming, great location, great service Clean, charming ho
tel located near the wonderful Rue Cler, a typical Fr"| __truncated__ "Intimate hotel with much to offer! I staye
d for 8 nights in late May 2022 at the Hotel de la Motte Picquet.  My"| __truncated__ ...
##  $ TripType : chr  "Trip type: Travelled with family" "Trip type: Travelled as a couple" "Trip type: Travelled
with friends" "Trip type: Travelled with family" ...
##  $ HotelName: chr  "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" ...
##  $ City     : chr  "There are more places to choose from in the Paris area." "There are more places to choose
from in the Paris area." "There are more places to choose from in the Paris area." "There are more places to choo
se from in the Paris area." ...
```

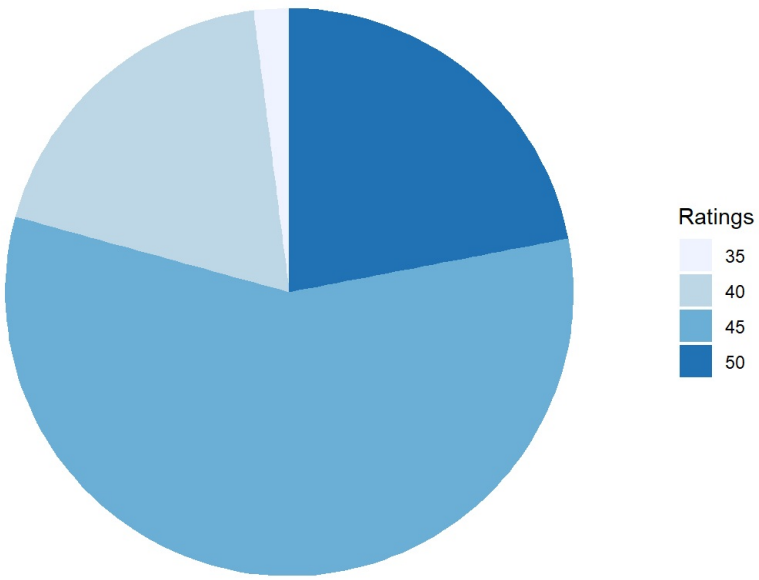
```
# Naturally Factor Variables
factor_vars = c("TripType")
df[factor_vars] = lapply(df[factor_vars], as.factor)
str(df)
```

```
## 'data.frame':    2230 obs. of  6 variables:
##  $ Name     : chr  "Summernole96" "beatrix659" "williamsS293YZ" "LizMSK" ...
##  $ Rating   : chr  "45" "45" "50" "45" ...
##  $ Review   : chr  "Great hotel for your trip to Paris Such a wonderful place to stay on our first visit to Pa
ris as a family. Ever"| __truncated__ "Lovely staying in Paris We stayed in this gorgeous little hotel with my fi
ance for a weekend in Paris. It has a"| __truncated__ "Charming, great location, great service Clean, charming ho
tel located near the wonderful Rue Cler, a typical Fr"| __truncated__ "Intimate hotel with much to offer! I staye
d for 8 nights in late May 2022 at the Hotel de la Motte Picquet.  My"| __truncated__ ...
##  $ TripType : Factor w/ 5 levels "Trip type: Travelled as a couple",..: 4 1 5 4 4 1 4 4 5 1 ...
##  $ HotelName: chr  "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" ...
##  $ City     : chr  "There are more places to choose from in the Paris area." "There are more places to choose
from in the Paris area." "There are more places to choose from in the Paris area." "There are more places to choo
se from in the Paris area." ...
```

```
blank_theme <- theme_minimal() +
  theme(
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
  panel.border = element_blank(),
  panel.grid=element_blank(),
  axis.ticks = element_blank(),
  plot.title=element_text(size=14, face="bold")
  )

df$Rating = as.factor(df$Rating)
str(df$Rating)
```

```
##  Factor w/ 4 levels "35","40","45",..: 3 3 4 3 4 3 4 4 4 4 ...
```

```
df_rating = as.data.frame(table(df$Rating))
ggplot(df_rating, aes(x="", y=Freq, fill=Var1)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start=0) +
  scale_fill_brewer("Ratings") + blank_theme +
  theme(axis.text.x=element_blank())
```




```
# City
df$City = sub("There are more places to choose from in the ", "", df$City)
df$City = sub(" area.", "", df$City)
table(df$City)
```

```
##
## Paris
##  2230
```

EDA: Word Cloud

```
# Creating Corpus
corpus = Corpus(VectorSource(df$Review))
# Clean
clean = tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
```

```
clean = tm_map(clean, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(clean, content_transformer(tolower)):
## transformation drops documents
```

```
clean = tm_map(clean, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(clean, removeNumbers): transformation drops
## documents
```

```
clean = tm_map(clean, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(clean, stripWhitespace): transformation drops
## documents
```

```
clean = tm_map(clean, removeWords,
               c(stopwords("english"), stopwords("french"),
                 "paris", "hotel", "one", "two", "airport", "terminal",
                 "really", "just", "flight", "very", "quite", "rather",
                 "didnt", "youre", "westin", "marriot", "stay", "stayed",
                 "will", "also", "day", "time", "rooms"
                 ))
```

```
## Warning in tm_map.SimpleCorpus(clean, removeWords, c(stopwords("english"), :
## transformation drops documents
```

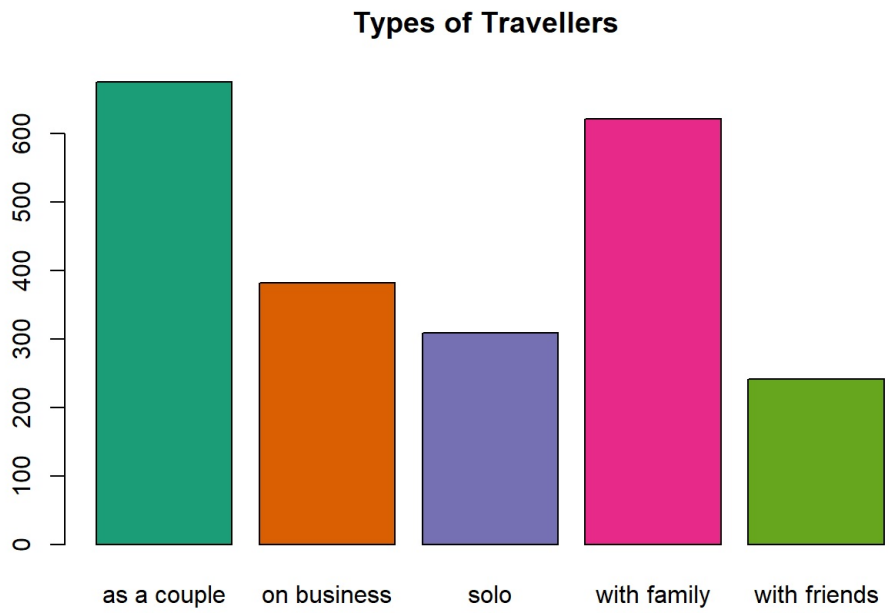
```
# Lementize corpus
word.cloud = lemmatize_words(clean)
saveRDS(word.cloud, "wordcloud_1.RDS")

# Word Cloud
wordcloud(word.cloud, random.order = F, max.words = 20,
          scale = c(5,1), colors = pallette)
```

EDA: Traveller Type

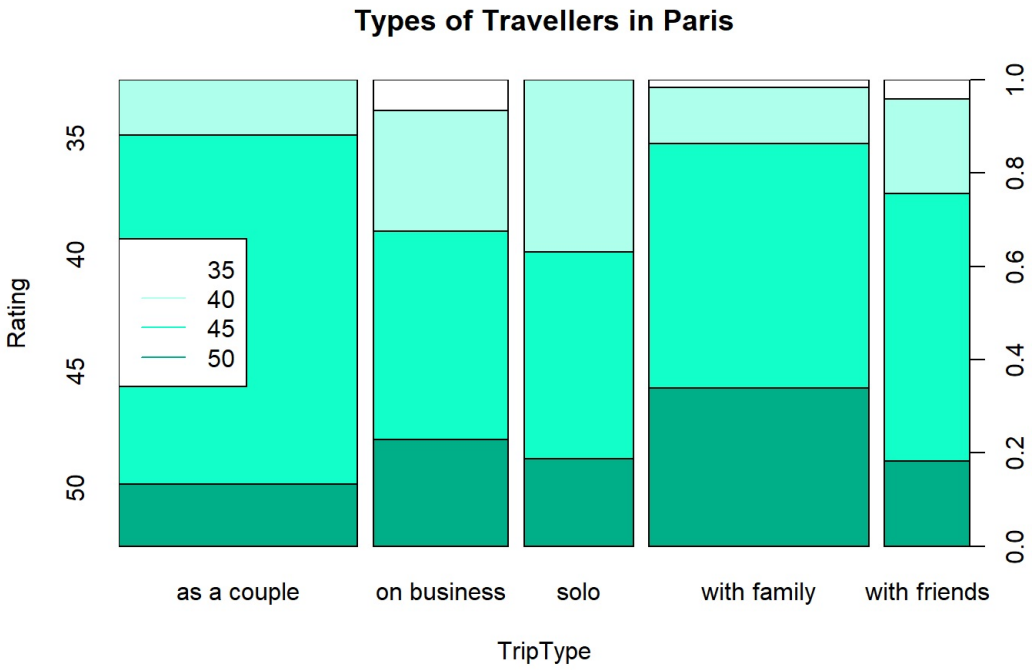
```
df$TripType = as.factor(sub('Trip type: Travelled ', '', df$TripType))
str(df)
```

```
## 'data.frame':    2230 obs. of  6 variables:
##  $ Name     : chr  "Summernole96" "beatrix659" "williamsS293YZ" "LizMSK" ...
##  $ Rating   : Factor w/ 4 levels "35","40","45",..: 3 3 4 3 4 3 4 4 4 4 ...
##  $ Review   : chr  "Great hotel for your trip to Paris Such a wonderful place to stay on our first visit to Pa
ris as a family. Ever"| __truncated__ "Lovely staying in Paris We stayed in this gorgeous little hotel with my fi
ance for a weekend in Paris. It has a"| __truncated__ "Charming, great location, great service Clean, charming ho
tel located near the wonderful Rue Cler, a typical Fr"| __truncated__ "Intimate hotel with much to offer! I staye
d for 8 nights in late May 2022 at the Hotel de la Motte Picquet.  My"| __truncated__ ...
##  $ TripType : Factor w/ 5 levels "as a couple",..: 4 1 5 4 4 1 4 4 5 1 ...
##  $ HotelName: chr  "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" "Hotel Motte Picquet" ...
##  $ City     : chr  "Paris" "Paris" "Paris" "Paris" ...
```

```
plot(df$TripType, main = "Types of Travellers ", col = pallette)
```




```
plot(Rating ~ TripType, data = df,
     col = c("#00af87", "#12ffc9", "#afffed", "#ffffff"),
     main = "Types of Travellers in Paris")
legend("left", legend = c("35", "40", "45", "50"),
       col = c("#ffffff", "#afffed", "#12ffc9", "#00af87"),
       lty=1, lwd=1)
```



Couples

```
df_couple = df[df$TripType == "as a couple",]
# Creating Corpus
corpus = Corpus(VectorSource(df_couple$Review))
# Clean
clean = tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
```

```
clean = tm_map(clean, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(clean, content_transformer(tolower)):
## transformation drops documents
```

```
clean = tm_map(clean, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(clean, removeNumbers): transformation drops
## documents
```

```
clean = tm_map(clean, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(clean, stripWhitespace): transformation drops
## documents
```

```
clean = tm_map(clean, removeWords,
               c(stopwords("english"), stopwords("french"),
                 "paris", "hotel", "one", "two", "airport", "terminal",
                 "really", "just", "flight", "very", "quite", "rather",
                 "didnt", "youre", "westin", "marriot", "stay", "stayed",
                 "will", "also", "day", "time", "rooms"
                 ))
```

```
## Warning in tm_map.SimpleCorpus(clean, removeWords, c(stopwords("english"), :
## transformation drops documents
```

```
# Lementize corpus
word.cloud = lemmatize_words(clean)
saveRDS(word.cloud, "wordcloud_1.RDS")

# Word Cloud
wordcloud(word.cloud, random.order = F, max.words = 10,
          scale = c(5,1), colors = pallette, main = "M")
```

Business

```
df_business = df[df$TripType == "on business",]
# Creating Corpus
corpus = Corpus(VectorSource(df_business$Review))
# Clean
clean = tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
```

```
clean = tm_map(clean, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(clean, content_transformer(tolower)):
## transformation drops documents
```

```
clean = tm_map(clean, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(clean, removeNumbers): transformation drops
## documents
```

```
clean = tm_map(clean, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(clean, stripWhitespace): transformation drops
## documents
```

```
clean = tm_map(clean, removeWords,
               c(stopwords("english"), stopwords("french"),
                 "paris", "hotel", "one", "two", "airport", "terminal",
                 "really", "just", "flight", "very", "quite", "rather",
                 "didnt", "youre", "westin", "marriot", "stay", "stayed",
                 "will", "also", "day", "time", "rooms"
                 ))
```

```
## Warning in tm_map.SimpleCorpus(clean, removeWords, c(stopwords("english"), :
## transformation drops documents
```

```
# Lementize corpus
word.cloud = lemmatize_words(clean)
saveRDS(word.cloud, "wordcloud_1.RDS")

# Word Cloud
wordcloud(word.cloud, random.order = F, max.words = 10,
          scale = c(5,1), colors = pallette, main = "M")
```

Solo

```
df_solo = df[df$TripType == "solo",]
# Creating Corpus
corpus = Corpus(VectorSource(df_solo$Review))
# Clean
clean = tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
```

```
clean = tm_map(clean, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(clean, content_transformer(tolower)):
## transformation drops documents
```

```
clean = tm_map(clean, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(clean, removeNumbers): transformation drops
## documents
```

```
clean = tm_map(clean, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(clean, stripWhitespace): transformation drops
## documents
```

```
clean = tm_map(clean, removeWords,
               c(stopwords("english"), stopwords("french"),
                 "paris", "hotel", "one", "two", "airport", "terminal",
                 "really", "just", "flight", "very", "quite", "rather",
                 "didnt", "youre", "westin", "marriot", "stay", "stayed",
                 "will", "also", "day", "time", "rooms"
                 ))
```

```
## Warning in tm_map.SimpleCorpus(clean, removeWords, c(stopwords("english"), :
## transformation drops documents
```

```
# Lementize corpus
word.cloud = lemmatize_words(clean)
saveRDS(word.cloud, "wordcloud_1.RDS")

# Word Cloud
wordcloud(word.cloud, random.order = F, max.words = 10,
          scale = c(5,1), colors = pallette, main = "M")
```

Family

```
df_family = df[df$TripType == "with family",]
# Creating Corpus
corpus = Corpus(VectorSource(df_family$Review))
# Clean
clean = tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
```

```
clean = tm_map(clean, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(clean, content_transformer(tolower)):
## transformation drops documents
```

```
clean = tm_map(clean, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(clean, removeNumbers): transformation drops
## documents
```

```
clean = tm_map(clean, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(clean, stripWhitespace): transformation drops
## documents
```

```
clean = tm_map(clean, removeWords,
               c(stopwords("english"), stopwords("french"),
                 "paris", "hotel", "one", "two", "airport", "terminal",
                 "really", "just", "flight", "very", "quite", "rather",
                 "didnt", "youre", "westin", "marriot", "stay", "stayed",
                 "will", "also", "day", "time", "rooms"
                 ))
```

```
## Warning in tm_map.SimpleCorpus(clean, removeWords, c(stopwords("english"), :
## transformation drops documents
```

```
# Lementize corpus
word.cloud = lemmatize_words(clean)
saveRDS(word.cloud, "wordcloud_1.RDS")

# Word Cloud
wordcloud(word.cloud, random.order = F, max.words = 10,
          scale = c(5,1), colors = pallette, main = "M")
```

Friends

```
df_friends = df[df$TripType == "as a couple",]
# Creating Corpus
corpus = Corpus(VectorSource(df_friends$Review))
# Clean
clean = tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
```

```
clean = tm_map(clean, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(clean, content_transformer(tolower)):
## transformation drops documents
```

```
clean = tm_map(clean, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(clean, removeNumbers): transformation drops
## documents
```

```
clean = tm_map(clean, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(clean, stripWhitespace): transformation drops
## documents
```

```
clean = tm_map(clean, removeWords,
               c(stopwords("english"), stopwords("french"),
                 "paris", "hotel", "one", "two", "airport", "terminal",
                 "really", "just", "flight", "very", "quite", "rather",
                 "didnt", "youre", "westin", "marriot", "stay", "stayed",
                 "will", "also", "day", "time", "rooms"
                 ))
```

```
## Warning in tm_map.SimpleCorpus(clean, removeWords, c(stopwords("english"), :
## transformation drops documents
```

```
# Lementize corpus
word.cloud = lemmatize_words(clean)
saveRDS(word.cloud, "wordcloud_1.RDS")

# Word Cloud
wordcloud(word.cloud, random.order = F, max.words = 10,
          scale = c(5,1), colors = pallette, main = "M")
```