

# R markdown\_MarieLaureJOVIAL\_EnronProject

2024-11-07

## Loading data and library

```
library(ggplot2)
library(dplyr)
library(bslib)
library(lubridate)
library(ggiraph)
library(stringr)
library(tidyr)
load('C:\\Users\\MarieLaureJovial\\Downloads\\R BIG DATA PROCESSING\\Enron.Rdata')
```

## Quick pre exploration of the data

```
## [1] " Number of messages: 252759"
```

```
## [1] " Number of person in employeeelist: 149"
```

```
## [1] " Number of recipients: 2064442"
```

Based on the second question in the exam instruction, we have a look at the different existent status and their distribution

```
table(employeeelist$status)
```

```
##
##          CEO          Director      Employee  In House Lawyer
##           4           14           41           1
##      Manager Managing Director      N/A      President
##           14           3           31           4
##      Trader   Vice President
##           13           23
```

A total of 252759 messages were sent to 2064442 recipients.

The company has registered 149 person under 11 different status.

To avoid any confusion with employee status, from now on we will refer to these 149 company employees using the term user.

We observe that some messages was sent with recipient directly or in a hidden manner.

```
recipientinfo %>% group_by(rtype) %>% summarise(ComType = n())
```

```
## # A tibble: 3 × 2
##   rtype ComType
##   <fct>   <int>
## 1 BCC     253713
## 2 CC      253735
## 3 TO      1556994
```

Recipients mostly received messages that were meant for them. An interesting point might be analyzing if the number of hidden recipient increased during tense period for the company.

## OBJECTIVE 1: ANALYZE THE EMAIL SENT BY USERS (USER FOCUSED)

Cleaning and making some transformation to the data

```
#Replacing the null status in employeeelist by N/A
employeeelist$status[is.na(employeeelist$status)] <- "N/A"

#Making sure that the date column is explorable for future visuals
message$date <- as.Date(trimws(ymd(message$date)))

#Adding a fullName column to employeeelist
employeeelistFN <- mutate(employeeelist, fullName = paste(employeeelist$firstName, employeeelist$lastName))
```

## RELEVANT QUESTION TO OBJECTIVE 1

**-How many messages were sent? how and by who ?**

**-Is there a group who is more active than another ?**

**Transforming employeeelist table :**

To answer these questions we must join the employeeelistFN table with the message table.

Given each employee can have up to 4 email, the join must be performed for each email.

```
#Join of employeeelist table and message table on Email_id
EmployeeelistFN1 <-
employeeelistFN %>%
select(eid,Email_id,status,fullName)%>%
inner_join(message, by=c("Email_id"="sender"))
dim(EmployeeelistFN1)
```

```
## [1] 104766      8
```

```
#Join of employeeelist table and message table on Email2
EmployeeelistFN2 <-
employeeelistFN %>%
filter(!is.na(Email2) & Email2 != "")%>%
select(eid,Email2,status,fullName)%>%
inner_join(message, by=c("Email2"="sender"))
dim(EmployeeelistFN2)
```

```
## [1] 0 8
```

```
#Join of employeelist table and message table on Email3
EmployeeelistFN3 <-
employeeelistFN %>%
filter(!is.na(Email3) & Email3 != "")%>%
select(eid,Email3,status,fullName)%>%
inner_join(message, by=c("Email3"="sender"))
dim(EmployeeelistFN3)
```

```
## [1] 1170      8
```

```
#Join of employeelist table and message table on Email4
EmployeeelistFN4 <-
employeeelistFN %>%
filter(!is.na(Email4) & Email4 != "")%>%
select(eid,Email4,status,fullName)%>%
inner_join(message, by=c("Email4"="sender"))
dim(EmployeeelistFN4)
```

```
## [1] 0 8
```

```
#Appending all 4 data
EmployeeMessageSent <- bind_rows(EmployeeelistFN1, EmployeeelistFN2, EmployeeelistFN3, EmployeeelistFN4)
```

#### COMMENT:

104766 emails were sent with Email\_id and 1170 emails were sent with Email\_3 which makes a **total of 105936 emails**. This also means that only **41.9% of the 252759 messages were sent by users**. Note that emails were only emitted from Email\_id and Email3.

## INVESTIGATING THE NUMBER OF EMAIL SENT BY USER

Who are the most active senders ?

```
#Calculating the amount of message sent by each user
MsgByEid <- EmployeeMessageSent %>% group_by(eid, fullName, status) %>% summarise(TotalMessageSent =
n()) %>% arrange(desc(TotalMessageSent))
MsgByEid
```

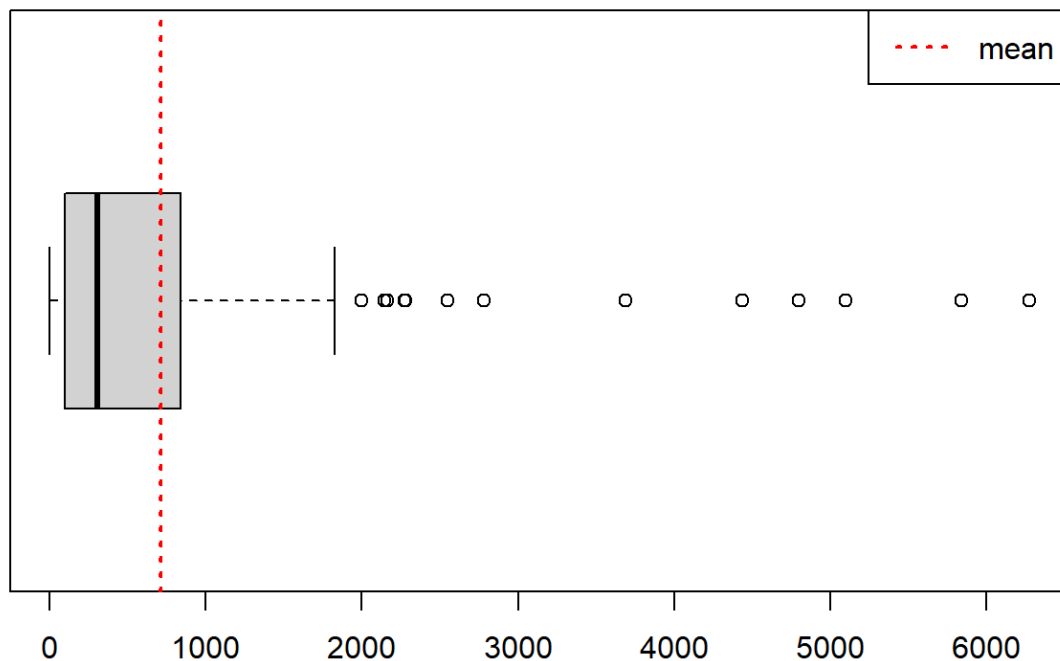
```
## # A tibble: 149 × 4
## # Groups:   eid, fullName [149]
##      eid fullName      status      TotalMessageSent
##    <dbl> <chr>      <fct>          <int>
##  1    73 Jeff Dasovich Employee          6273
##  2   109 Vince Kaminski Manager           5838
##  3    52 Kay Mann      Employee          5100
##  4    67 Sara Shackleton N/A              4797
##  5    48 Tana Jones     N/A              4437
##  6   114 Chris Germany  Employee          3686
##  7    61 Matthew Lenhart Employee          2784
##  8    76 Debra Perlingiere Employee          2549
##  9    51 Steven Kean     Vice President    2278
## 10   75 Gerald Nemec     N/A              2275
## # i 139 more rows
```

```
#Calculating the amount of message sent by an average user
MeanofMsgByEid <- round(mean(MsgByEid$TotalMessageSent),0)
print(paste("In average a user sends ", MeanofMsgByEid, "messages"))
```

```
## [1] "In average a user sends 711 messages"
```

```
#Visualising the distribution of message sent by user
boxplot(MsgByEid$TotalMessageSent, horizontal = TRUE)
title("Message Sent Distribution")
abline(v=MeanofMsgByEid, lty=3, lwd=2, col="red")
legend("topright", legend = c('mean'), lty = 3, lwd = 2, col = 'red')
```

## Message Sent Distribution



### COMMENT:

The most active sender is the Employee Jeff Dasovich with 6273 emails sent.

The less active sender is the N/A Steven Merris with 1 email sent.

In average a user sends 711 emails.

However, 50% of the users have sent between 101 and 843 messages (Q1-Q3).

The average of message sent being so close to the 3rd quantile and so far from the average shows that some user are sending an extreme amount of messages which is leading to such a difference between the median and the mean.

*#Defining if the top active emailers where still there during tense period of the company*

```
library(lubridate)
EmployeeMessageSent$date <- ymd(EmployeeMessageSent$date)
LastSent <- EmployeeMessageSent %>% select (fullName,date) %>% group_by(fullName) %>% mutate (LatestEmail= max(date)) %>% distinct(fullName, LatestEmail, .keep_all = FALSE)

TopSender_LatestEmailSent<-merge(MsgByEid,LastSent) %>% arrange(desc(TotalMessageSent))
head(TopSender_LatestEmailSent,15)
```

##	fullName	eid	status	TotalMessageSent	LatestEmail
## 1	Jeff Dasovich	73	Employee	6273	2002-09-22
## 2	Vince Kaminski	109	Manager	5838	2002-01-30
## 3	Kay Mann	52	Employee	5100	2002-05-28
## 4	Sara Shackleton	67	N/A	4797	2002-03-25
## 5	Tana Jones	48	N/A	4437	2002-02-08
## 6	Chris Germany	114	Employee	3686	2002-06-25
## 7	Matthew Lenhart	61	Employee	2784	2002-02-05
## 8	Debra Perlingiere	76	Employee	2549	2002-03-25
## 9	Steven Kean	51	Vice President	2278	2001-12-27
## 10	Gerald Nemec	75	N/A	2275	2002-02-21
## 11	James Steffes	23	Vice President	2161	2002-02-03
## 12	Mark Taylor	6	Employee	2146	2001-11-27
## 13	Richard Sanders	14	Vice President	1998	2002-02-19
## 14	Carol Clair	147	Vice President	1827	2001-06-11
## 15	Susan Scott	49	N/A	1823	2002-05-22

## INVESTIGATING THE POSSIBLE INFLUENCE OF STATUS ON EMAILS SENT.

### Overview by status

```
#Using the table function to summarize the message sent by status
table(EmployeeMessageSent$status)
```

##	CEO	Director	Employee	In House Lawyer
##	1388	3005	37267	763
##	Manager	Managing Director	N/A	President
##	13447	1176	24313	3619
##	Trader	Vice President		
##	3800	17158		

### Which groups are the most active ?

```
#Defining the quantity of email sent by status in absolute number and in percentage.
```

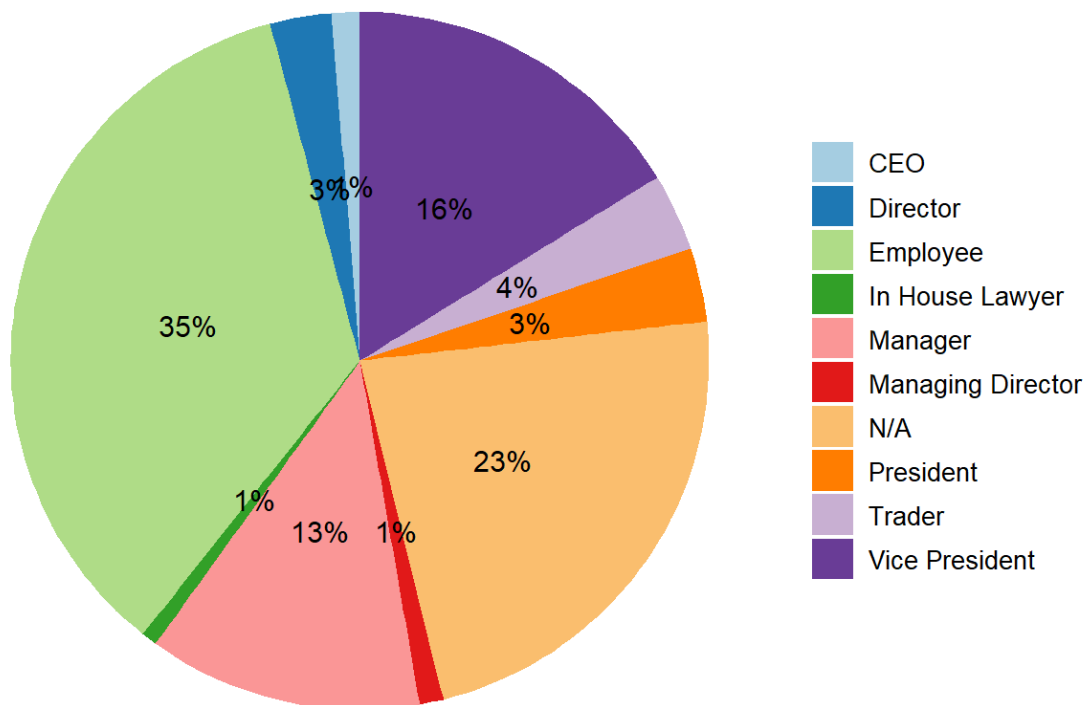
```
MsgByStatus <- EmployeeMessageSent %>% group_by(status) %>% summarise(TotalSent = n()) %>% arrange(desc(TotalSent))%>% mutate(TotalSentPercentage= round(TotalSent/sum(TotalSent)*100))
MsgByStatus
```

```
## # A tibble: 10 × 3
##   status      TotalSent TotalSentPercentage
##   <fct>      <int>          <dbl>
## 1 Employee      37267             35
## 2 N/A           24313             23
## 3 Vice President 17158             16
## 4 Manager       13447             13
## 5 Trader        3800              4
## 6 President     3619              3
## 7 Director      3005              3
## 8 CEO           1388              1
## 9 Managing Director 1176              1
## 10 In House Lawyer 763              1
```

*#Representing the repartition in % in a pie chart*

```
ggplot(data = MsgByStatus, aes(x = '', y = TotalSent, fill = status)) +
  geom_bar(stat = 'identity', width = 1) + # Create bars to form a pie
  geom_text(aes(label = paste0(TotalSentPercentage, '%')),
            position = position_stack(vjust = 0.5), color = 'black', size = 4) + # Add percentage labels
  coord_polar(theta = 'y') + # Turn the bar chart into a pie chart
  theme_void() + # Remove background
  scale_fill_brewer(palette = "Paired") + # Choose a color palette for better readability
  theme(legend.title = element_blank(), # Remove Legend title for clarity
        legend.text = element_text(size = 10))+ # Adjust legend text size
  ggtitle("Distribution of Email Sent ")
```

Distribution of Email Sent



**COMMENT:** As Employees, N/A's and Vice President user status are the most common it is with little surprise that, 35% messages were sent by employees, 23% of messages were sent by N/A's (23%) followed by 16% sent by Vice Presidents. Indeed out of the 149 users, 41 are Employees, 31 are N/A's and 23 are Vice President (see exploration part)

However we can further investigate if this is the only explanation

## Further exploration of status impact with basic statistic concepts.

```
MsgBySender <- EmployeeMessageSent %>% group_by(eid, fullName, status)%>% summarise(TotalSent = n())
MsgBySender_CEO <- select(filter(MsgBySender, status=="CEO"),status,TotalSent)
MsgBySender_Director <- select(filter(MsgBySender, status=="Director"),status,TotalSent)
MsgBySender_Employee <- select(filter(MsgBySender, status=="Employee"),status,TotalSent)
MsgBySender_IHL <- select(filter(MsgBySender, status=="In House Lawyer"),status,TotalSent)
MsgBySender_Manager <- select(filter(MsgBySender, status=="Manager"),status,TotalSent)
MsgBySender_MDirector <- select(filter(MsgBySender, status=="Managing Director"),status,TotalSent)
MsgBySender_NA <- select(filter(MsgBySender, status=="N/A"),status,TotalSent)
MsgBySender_President <- select(filter(MsgBySender, status=="President"),status,TotalSent)
MsgBySender_Trader <- select(filter(MsgBySender, status=="Trader"),status,TotalSent)
MsgBySender_VicePresident <- select(filter(MsgBySender, status=="Vice President"),status,TotalSent)
summary(MsgBySender_CEO)
```

##	eid	fullName	status	TotalSent
##	Min. : 46.00	Length:4	CEO	:4 Min. : 30.00
##	1st Qu.: 51.25	Class :character	Director	:0 1st Qu.: 59.25
##	Median : 90.00	Mode :character	Employee	:0 Median :236.00
##	Mean : 88.75		In House Lawyer	:0 Mean :347.00
##	3rd Qu.:127.50		Manager	:0 3rd Qu.:523.75
##	Max. :129.00		Managing Director:	0 Max. :886.00
##			(Other)	:0

```
summary(MsgBySender_Director)
```

##	eid	fullName	status	TotalSent
##	Min. : 1.00	Length:14	Director	:14 Min. : 11.00
##	1st Qu.: 22.25	Class :character	CEO	: 0 1st Qu.: 31.25
##	Median : 99.00	Mode :character	Employee	: 0 Median : 112.00
##	Mean : 73.79		In House Lawyer	: 0 Mean : 214.64
##	3rd Qu.:105.50		Manager	: 0 3rd Qu.: 205.75
##	Max. :130.00		Managing Director:	0 Max. :1116.00
##			(Other)	: 0

```
summary(MsgBySender_Employee)
```



```
##      eid      fullName      status      TotalSent
## Min.   : 3.00   Length:41      Employee      :41   Min.    : 2
## 1st Qu.: 25.00   Class :character   CEO          : 0   1st Qu.: 100
## Median : 64.00   Mode  :character   Director     : 0   Median : 319
## Mean   : 66.15                In House Lawyer : 0   Mean   : 909
## 3rd Qu.:100.00                Manager        : 0   3rd Qu.:1110
## Max.   :149.00                Managing Director: 0   Max.    :6273
##                                (Other)         : 0
```

```
summary(MsgBySender_IHL)
```

```
##      eid      fullName      status      TotalSent
## Min.   :74   Length:1      In House Lawyer :1   Min.    :763
## 1st Qu.:74   Class :character   CEO            :0   1st Qu.:763
## Median :74   Mode  :character   Director       :0   Median :763
## Mean   :74                Employee       :0   Mean   :763
## 3rd Qu.:74                Manager        :0   3rd Qu.:763
## Max.   :74                Managing Director:0   Max.    :763
##                                (Other)         :0
```

```
summary(MsgBySender_Manager)
```

```
##      eid      fullName      status      TotalSent
## Min.   : 4.0   Length:14      Manager        :14   Min.    : 18.0
## 1st Qu.: 28.5   Class :character   CEO            : 0   1st Qu.: 216.2
## Median : 99.0   Mode  :character   Director       : 0   Median : 556.5
## Mean   : 74.0                Employee       : 0   Mean   : 960.5
## 3rd Qu.:115.5                In House Lawyer : 0   3rd Qu.: 969.5
## Max.   :123.0                Managing Director: 0   Max.    :5838.0
##                                (Other)         : 0
```

```
summary(MsgBySender_MDirector)
```

```
##      eid      fullName      status      TotalSent
## Min.   : 65.0   Length:3      Managing Director:3   Min.    : 9.0
## 1st Qu.: 87.5   Class :character   CEO              :0   1st Qu.: 28.0
## Median :110.0   Mode  :character   Director         :0   Median : 47.0
## Mean   :104.3                Employee        :0   Mean   : 392.0
## 3rd Qu.:124.0                In House Lawyer :0   3rd Qu.: 583.5
## Max.   :138.0                Manager         :0   Max.    :1120.0
##                                (Other)         :0
```

```
summary(MsgBySender_NA)
```

```
##      eid      fullName      status      TotalSent
## Min.   : 11.00 Length:32      N/A      :32 Min.    : 1.0
## 1st Qu.: 49.75 Class :character CEO      : 0 1st Qu.: 123.5
## Median : 77.00 Mode  :character Director  : 0 Median : 307.5
## Mean   : 80.91      Employee   : 0 Mean   : 759.8
## 3rd Qu.:107.50      In House Lawyer: 0 3rd Qu.: 835.5
## Max.   :150.00      Manager    : 0 Max.    :4797.0
##                                     (Other)   : 0
```

```
summary(MsgBySender_President)
```

```
##      eid      fullName      status      TotalSent
## Min.   : 36.00 Length:4      President :4 Min.    : 223.0
## 1st Qu.: 49.50 Class :character CEO      :0 1st Qu.: 411.2
## Median : 60.00 Mode  :character Director  :0 Median : 946.0
## Mean   : 65.75      Employee   :0 Mean   : 904.8
## 3rd Qu.: 76.25      In House Lawyer:0 3rd Qu.:1439.5
## Max.   :107.00      Manager    :0 Max.    :1504.0
##                                     (Other)   :0
```

```
summary(MsgBySender_Trader)
```

```
##      eid      fullName      status      TotalSent
## Min.   : 22.00 Length:13      Trader    :13 Min.    : 24.0
## 1st Qu.: 58.00 Class :character CEO      : 0 1st Qu.: 73.0
## Median : 91.00 Mode  :character Director  : 0 Median : 193.0
## Mean   : 91.77      Employee   : 0 Mean   : 292.3
## 3rd Qu.:131.00      In House Lawyer: 0 3rd Qu.: 235.0
## Max.   :134.00      Manager    : 0 Max.    :1644.0
##                                     (Other)   : 0
```

```
summary(MsgBySender_VicePresident)
```

```
##      eid      fullName      status      TotalSent
## Min.   : 14.0 Length:23      Vice President :23 Min.    : 87
## 1st Qu.: 37.5 Class :character CEO      : 0 1st Qu.: 283
## Median : 68.0 Mode  :character Director  : 0 Median : 502
## Mean   : 70.3      Employee   : 0 Mean   : 746
## 3rd Qu.: 94.0      In House Lawyer: 0 3rd Qu.: 912
## Max.   :147.0      Manager    : 0 Max.    :2278
##                                     (Other)   : 0
```

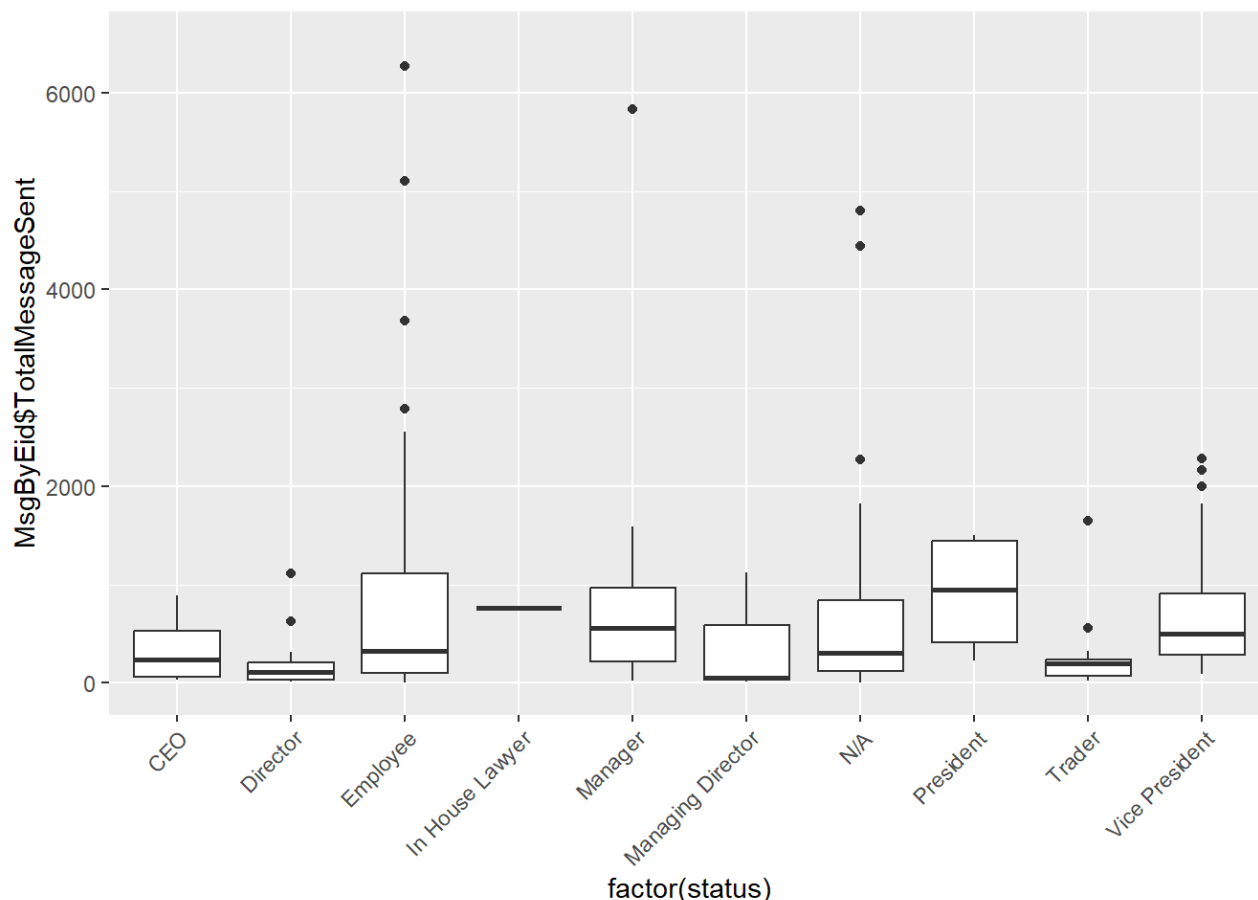
**COMMENT:** Previously we found out that the average user was sending 710 messages. From exploring the distribution of message by status, we learn the following: In addition of being the most numerous group, CEO's , N/A's and Employees also sent more messages in averages that the rest of the status which also explain the fact that they the top status sending the most messages

CEO and Presidents status seems to be the status with the most centered distribution and non excessive outliers ( raisonnable maximum)

For other status, we observe a big dispersion in the data. The means can be way superior to the median ( especially for Employees, Manahers, Managing Director' and N/A's ) explain by the presence of many users in these status being more active ( larger gap between Q3 and max)

We can illustrate this comment using a box plot by status

```
#Visualising previous comment with Boxplot by status
ggplot(data=MsgByEid)+ geom_boxplot(aes(x=factor(status),y=MsgByEid$TotalMessageSent))+
  ylim(0,6500) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



## OBJECTIVE 2: ANALYZE THE EMAIL RECIPIENTS (USER FOCUSED)

### Overview of the received emails (user AND non-user)

```
NB_Recipient <- nrow(recipientinfo)
NB_UniqueRecipient <- length(unique(recipientinfo$rvalue))
NB_UniqueMessages <- length(unique(recipientinfo$mid))
print(paste("Number of recipients:", NB_Recipient))
```

```
## [1] "Number of recipients: 2064442"
```

```
print(paste("Number of unique recipients:", NB_UniqueRecipient))
```

```
## [1] "Number of unique recipients: 68084"
```

```
print(paste("Number of unique messages:", NB_UniqueMessages))
```

```
## [1] "Number of unique messages: 252759"
```

### COMMENT:

Quantity of all recipients/messages received (by user and non user): 2064442.

Number of unique recipient (user and non user): 68084.

Number of unique message received by all (user and non-user): 252759.

The messages were sent differently to the recipients. It could be directly (TO), or indirectly (CC) or indirectly and hidden (BCC). Below the repartition:

```
RtypeDistrib <- table(recipientinfo$rtype)
RtypeDistrib
```

```
##
##      BCC      CC      TO
## 253713 253735 1556994
```

## INVESTIGATING EMAILS RECEIVED BY USERS (user only)

As saw previously , the users have several email addresses. To facilitate the research of email received independently of the email address, we built and AllUser\_wEmail table.

```
# DATASET AllUser_wEmail: a table with the employees and all their email adresses in one unique column (FullName, eid, status, Email). This can facilitate our exploration of the recipient.
```

```
EmployeePartID <- employeeelistFN %>% select (fullName, eid, status, Email_id) %>% rename(Email=Email_id) %>% mutate(EmailIndex = "Email_id")
```

```
EmployeePart2 <- employeeelistFN %>% filter(!is.na(Email2) & Email2 != "") %>% select(fullName,eid,status, Email2) %>% rename(Email=Email2) %>% mutate(EmailIndex = "Email2")
```

```
EmployeePart3 <- employeeelistFN %>% filter(!is.na(Email3) & Email3 != "") %>% select (fullName, eid, status, Email3) %>% rename(Email=Email3) %>% mutate(EmailIndex = "Email3")
```

```
EmployeePart4 <- employeeelistFN %>% filter(!is.na(Email4) & Email4 != "") %>% select (fullName, eid, status, Email4) %>% rename(Email=Email4) %>% mutate(EmailIndex = "Email4")
```

```
AllUser_wEmail <- bind_rows(EmployeePartID, EmployeePart2, EmployeePart3, EmployeePart4)
head(AllUser_wEmail,5)
```

##	fullName	eid	status	Email	EmailIndex
## 1	Marie Heard	13	N/A	marie.heard@enron.com	Email_id
## 2	Mark Taylor	6	Employee	mark.e.taylor@enron.com	Email_id
## 3	Lindy Donoho	19	Employee	lindy.donoho@enron.com	Email_id
## 4	Lisa Gang	115	N/A	lisa.gang@enron.com	Email_id
## 5	Jeffrey Skilling	129	CEO	jeff.skilling@enron.com	Email_id

## RELEVANT QUESTIONS LINK TO OBJECTIVE 2:

- How many emails did users received ?
- Are there user that received no emails at all ?
- Which user received the most/ the least email?
- Are there emails that were communicated to all the users and what were they about ?

*#Merging the recipientinfo table with the users*

```
UserRecipient <- merge(recipientinfo, AllUser_wEmail, by.x = "rvalue", by.y = "Email")
```

### COMMENT:

The users appeared as recipient 363616 times out of the 2064442 received email meaning 18% of the time.

*#Table of amount of message received by user*

```
MsgByUserRecipient<- UserRecipient %>% group_by(eid, fullName, status)%>% summarise(TotalReceived = n
(), .groups = "keep")%>% arrange(desc(TotalReceived))
```

### COMMENT:

The user who received the most messages is the Employee Jeff Dasovich with 11137 messages received.

The user who received the less messages is the Employee Mary Fischer with 140 messages received.

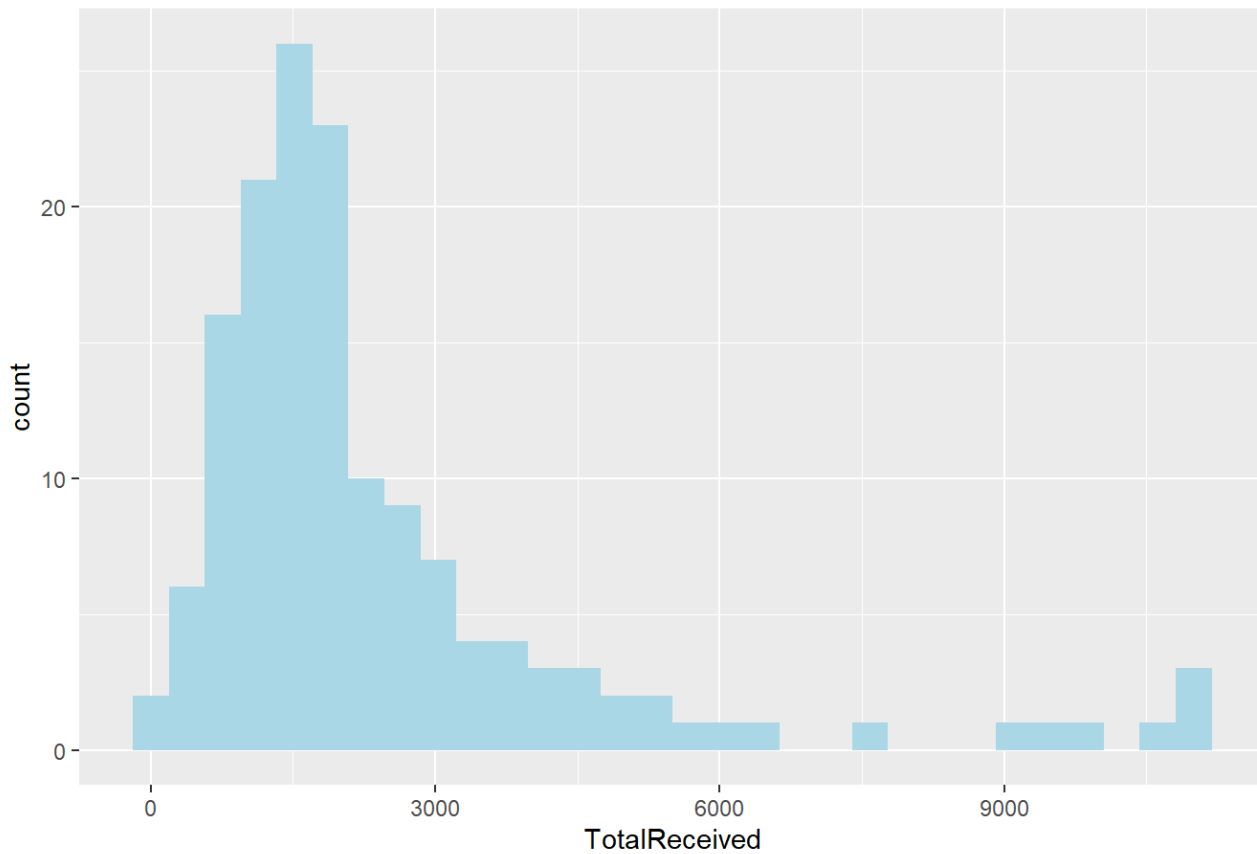
50% of the users receive between NA and 2725 messages.

## How many messages is commonly received ?

*# Plotting amount of message received by user in a histogram*

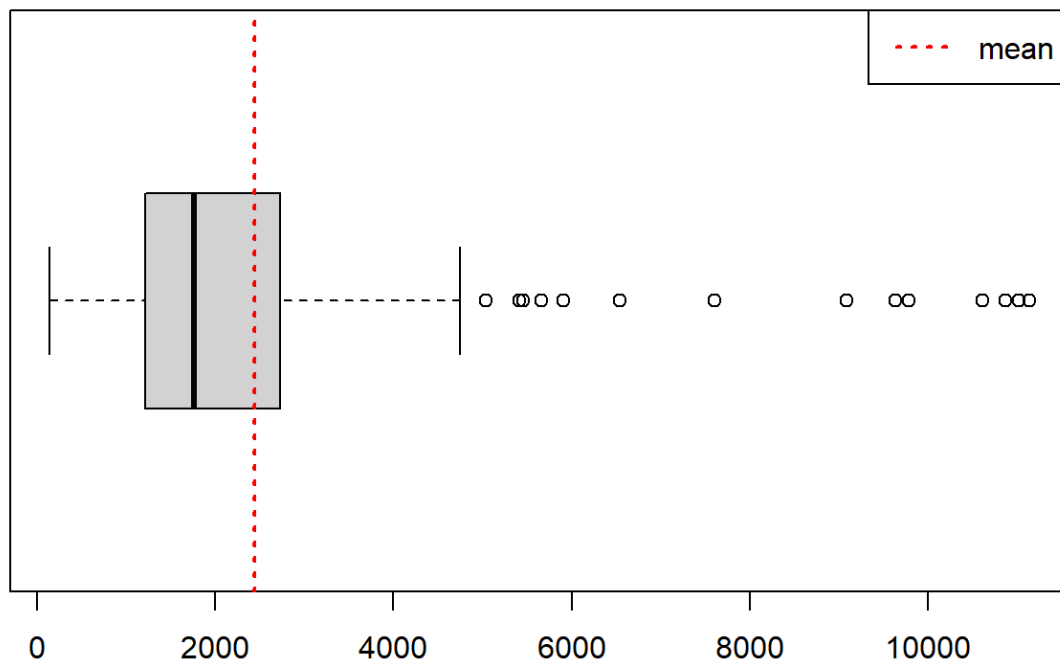
```
ggplot(data=MsgByUserRecipient)+
  geom_histogram(aes(x=TotalReceived), fill='lightblue')+
  ggtitle("Message Received frequency")
```

## Message Received frequency



```
# Understanding the distribution of amount of message received by user in a box plot
boxplot(MsgByUserRecipient$TotalReceived, horizontal = TRUE)
title("Message Received Distribution")
abline(v= round(mean(MsgByUserRecipient$TotalReceived),0), lty=3, lwd=2, col="red")
legend("topright", legend = c('mean'), lty = 3, lwd = 2, col = 'red')
```

## Message Received Distribution



```
#Calculating the amount of message sent by an average user
MeanofMsgByUserRecipient <- round(mean(MsgByUserRecipient$TotalReceived),0)
print(paste("In average a user receives ", MeanofMsgByUserRecipient, "messages"))
```

```
## [1] "In average a user receives 2440 messages"
```

### COMMENT :

Users received between close to 0 to 9000 messages. A vast majority of users received around 1500 messages (1000-2000). However we notice a few users receiving more than 3000 messages up to 12 000 messages

The box plot show us indeed a median close 1500 but because of the user receiving between 3000 and 12 000 emails, the mean averages 2500.

Indeed an average user received 2440 messages.

Is there messages that were broadly communicated or communicated to the entire organization ?

```
#Table of the messages that were received by all the organization
ReceivedByAll <- UserRecipient %>% group_by(mid)%>% summarise(TotalUserRecipient = n()) %>% filter(TotalUserRecipient>=148)

#Messages that were the most received in the organisation
MidMostReceived <- UserRecipient %>%
  group_by(mid)%>%
  summarise(TotalUserRecipient = n()) %>%
  filter(TotalUserRecipient==max(TotalUserRecipient))

ReceivedByAll
```

```
## # A tibble: 0 × 2
## #   mid <dbl>, TotalUserRecipient <int>
```

```
MidMostReceived
```

```
## # A tibble: 31 × 2
##       mid TotalUserRecipient
##   <dbl>         <int>
## 1 252612             67
## 2 255638             67
## 3 257022             67
## 4 257612             67
## 5 258164             67
## 6 267566             67
## 7 269583             67
## 8 270669             67
## 9 271748             67
## 10 272716            67
## #   21 more rows
```

## COMMENT

0 messages was received by all user in the company. Indeed , no message appear to have be received by 148 person (149 user - 1sender).

The maximum of recipient for a unique message was 67 and it was the case for 31 messages.

## (suite) OTHER QUESTIONS ADDRESSED RELATED TO OBJECTIVE 2:

**-Is there a status group that receives more emails than others?**

**-Does the communication style varies according to status?**

```
#Table of number of message received by user by rtype
MsgRtypeByUserStatus<- UserRecipient %>%
  group_by(status, rtype)%>%
  summarise(TotalReceived = n())
MsgRtypeByUserStatus
```

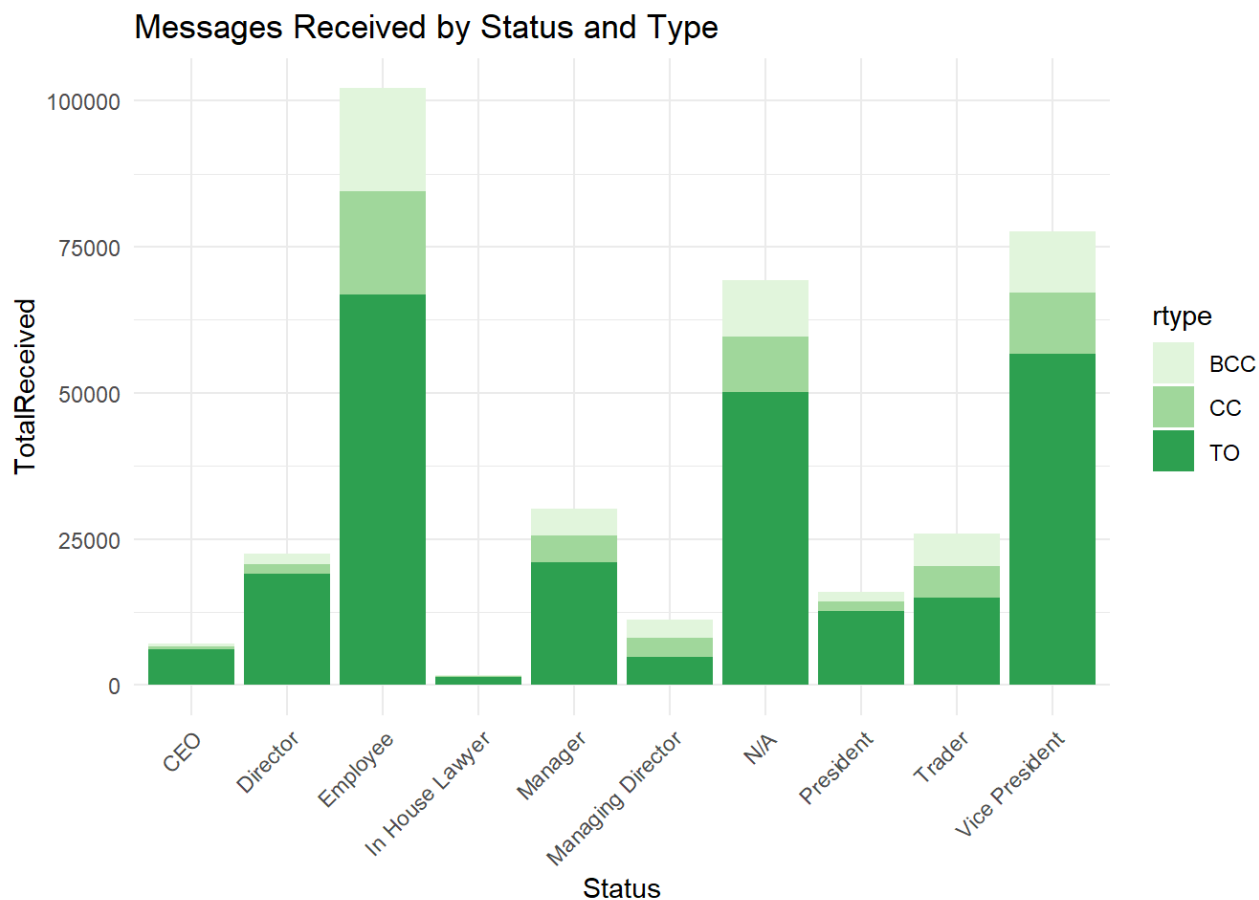


```
## # A tibble: 30 × 3
## # Groups:   status [10]
##   status      rtype TotalReceived
##   <fct>      <fct>      <int>
## 1 CEO        BCC          535
## 2 CEO        CC           535
## 3 CEO        TO        6050
## 4 Director   BCC        1740
## 5 Director   CC        1740
## 6 Director   TO       19002
## 7 Employee   BCC       17703
## 8 Employee   CC       17703
## 9 Employee   TO      66773
## 10 In House Lawyer BCC         186
## # i 20 more rows
```

**!It looks like all messages sent in CC was automatically sent in BCC or inversely**

Let's analyze the proportion represented by each type of rtype for every status

```
ggplot(MsgRtypeByUserStatus, aes(x = status, y = TotalReceived, fill = rtype)) +
  geom_bar(stat = "identity") +
  labs(title = "Messages Received by Status and Type", x = "Status", y = "TotalReceived") +
  theme_minimal() +
  scale_fill_brewer(palette = "Greens")+ theme(axis.text.x = element_text(angle = 45, hjust =
1))
```



**COMMENT :** Employees were the one that received the most email. But all status received mostly messages directly address to them.

However, we can notice that for the status that are C-level(President, CEO, Director), the indirect messages received represent a smaller portion of their total emails compared to other status.

This make sens has people higher in organization hierarchy are less involved in the day to day activities and are usually solicited only when really needed.

## What is the dynamic of messages received vs messages sent by status ?

```
#Preparing the table that will be use to answer the question

#Number message sent by status (in value and percentage) - renaming the table for consistency
MsgSentByStatus <- MsgByStatus

#Number message rceived by status (in value and percentage)
MsgReceibedByStatus<- UserRecipient %>%
  filter(rtype!="BCC")%>%
  group_by(status)%>%
  summarise(TotalReceived = n())%>%
  mutate(TotalReceivedPercentage= round(TotalReceived/sum(TotalReceived)*100))

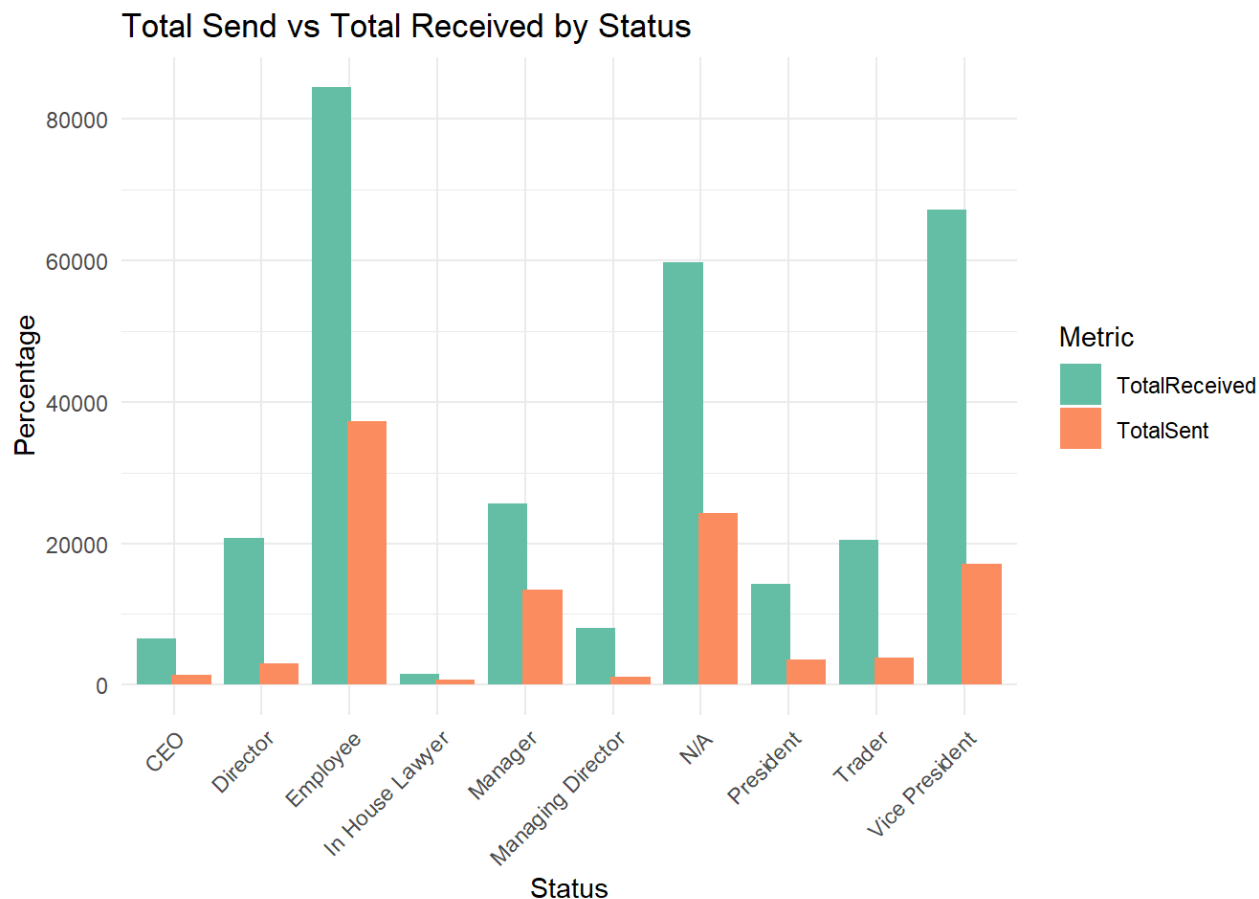
#Number user by status (in value and percentage)
EmployeeByStatus <- employeeelistFN %>%
  group_by(status)%>%
  summarise(TotalUser = n())%>%
  mutate(TotalUserPercentage= round(TotalUser/sum(TotalUser)*100))

#Merge of the 3 tables below
StatusTable <- MsgSentByStatus %>%
  inner_join(MsgReceibedByStatus, by = "status") %>%
  inner_join(EmployeeByStatus, by = "status")
head(StatusTable,2)
```

```
## # A tibble: 2 × 7
##   status   TotalSent TotalSentPercentage TotalReceived TotalReceivedPercentage
##   <fct>     <int>           <dbl>         <int>           <dbl>
## 1 Employee   37267             35         84476             27
## 2 N/A       24313             23         59686             19
## # i 2 more variables: TotalUser <int>, TotalUserPercentage <dbl>
```

```
# Reshape the data to Long format
StatusTable_long2 <- StatusTable %>%
  pivot_longer(
    cols = c(TotalReceived, TotalSent),
    names_to = "Metric",
    values_to = "Percentage"
  )

# Create the grouped bar plot with dodged bars
ggplot(StatusTable_long2, aes(x = factor(status), y = Percentage, fill = Metric)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(
    title = "Total Send vs Total Received by Status",
    x = "Status",
    y = "Percentage",
    fill = "Metric"
  ) +
  theme_minimal()+
  scale_fill_brewer(palette = "Set2")+ theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



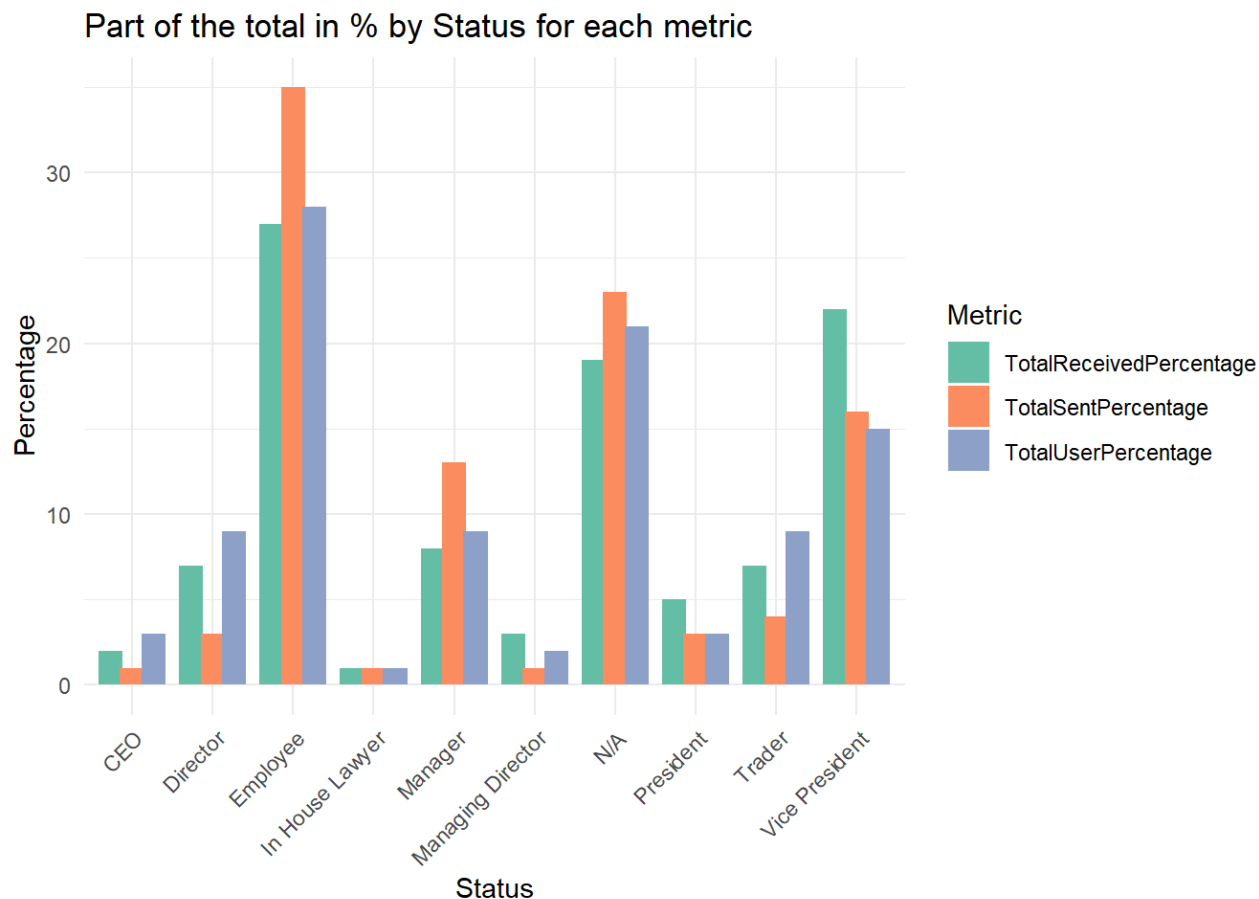
**COMMENT :**For all status there were more messages received than messages sent.

But as certain status are more numerous than other it could be interesting to express the same data taking this information into account.

Considering that other status are more numerous than other, we decide to express all metrics as a percentage of the total.

```
# Reshape the data to Long format
StatusTable_long <- StatusTable %>%
  pivot_longer(
    cols = c(TotalReceivedPercentage, TotalSentPercentage, TotalUserPercentage),
    names_to = "Metric",
    values_to = "Percentage"
  )

# Create the grouped bar plot with dodged bars
ggplot(StatusTable_long, aes(x = factor(status), y = Percentage, fill = Metric)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) + # Use position_dodge
  labs(
    title = "Part of the total in % by Status for each metric",
    x = "Status",
    y = "Percentage",
    fill = "Metric"
  ) +
  theme_minimal()+
  scale_fill_brewer(palette = "Set2")+ theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



**COMMENT:** Put under the same scale, the latter plot seems to reinforce our precedent findings.

Employees and N/A's, status that might be more involved in day-to-day operation and common/group projects seems to exchange more.

Vice President also have a high participation in exchanges but we observe that unlike the two precedent status, The Vice President received more emails than he sends. This could be explain by the more important CC and BCC messages he has been receiving compared to the other upper class status.

For the trader status, the communication seems limited. They represent close to 9% of the user but only 4% of the messages sent and 7% of the messages received. This could be explain by the company organization, if each trader has his own stock portfolio to manage then few exchanges is needed.

To conclude our objective wa can say that Employee ans N/A's seems to be the group the most actives both in term of sending ans receiving email. This plot shows a hierarchical organization and descending type of communication.

## OBJECTIVE 3: ANALYZE ALL EMAILS DYNAMIC OVER TIME AND SENDER - RECIPIENT RELATION

(general focus: user and non-user)

### RELEVANT INTERROGATIONS RELATED TO OBJECTIVE 3

- What was the dates /the periods on which the most emails were exchange?
- Use of specific email addresses or not overtime or during key periods
- Message overtime between status A and status B\_given the key date events
- Message overtime between NON(user)/ (NON)user : do user received email outside of their main email address?, how many of the exchanges were between users?
- How many times did the user sent email outside out the organization? Who sends and received the most email to/from non user
- How many time did the users received email coming from outside the organisation ? -Are email that goes out of the organization sent mostly from non primary email addresses?

#### GENERAL TABLE

Building a general table including all recipient, all senders and all messages with their subject and content as previous dataset and tables were user-focused.

```
#Building a general dataset by merging the 4 original data in the enron Rdata file

AllMessage_AllSender <- left_join(message,AllUser_wEmail, by = c("sender" = "Email"))

AllReceivers <- left_join(recipientinfo,AllUser_wEmail, by = c("rvalue" = "Email"))

AllMessage_AllSender_AllReceivers <- full_join(AllMessage_AllSender,AllReceivers, by = c("mid" = "mid")) %>% rename(eid_Emailer = eid.x, fullName_Emailer = fullName.x, status_Emailer = status.x, EmailIndex_Emailer =EmailIndex.x, eid_Recipient = eid.y,fullName_Recipient = fullName.y, status_Recipient = status.y, EmailIndex_Recipient =EmailIndex.y)

AllMessage_AllSender_AllReceivers_wContent <- full_join(AllMessage_AllSender_AllReceivers,referenceinfo, by = c("mid" = "mid")) %>% filter(year(date)>=1999 & year(date)<=2002)
```

Transforming the general table by adding column that could help with our analysis.

```
#Adding a column to indicate if the sender/receiver was a user or not
#Adding a column to identify the user status involved in an exchange
#Adding a column to help group data by month date if needed

CompleteDataset <- AllMessage_AllSender_AllReceivers_wContent %>%
  mutate(ExchangeCat = case_when(
    is.na(eid_Emailer) & is.na(eid_Recipient) ~ "NON user - NON user",
    !is.na(eid_Emailer) & is.na(eid_Recipient) ~ "user - NON user",
    is.na(eid_Emailer) & !is.na(eid_Recipient) ~ "NON user - user",
    !is.na(eid_Emailer) & !is.na(eid_Recipient) ~ "user - user"
  ))%>%
  mutate(ExchangeStatus = paste(status_Emailer,status_Recipient,sep = "-"))%>%
  mutate(month_year = floor_date(date, "month"))
```

Column name of the complete dataset

```
colnames(CompleteDataset)
```

```
## [1] "mid"           "sender"        "date"
## [4] "message_id"    "subject"       "fullName_Emailer"
## [7] "eid_Emailer"   "status_Emailer" "EmailIndex_Emailer"
## [10] "rid"           "rtype"         "rvalue"
## [13] "fullName_Recipient" "eid_Recipient" "status_Recipient"
## [16] "EmailIndex_Recipient" "rfid"         "reference"
## [19] "ExchangeCat"   "ExchangeStatus" "month_year"
```

## STUDYING TREND OF EMAIL OVER TIME

### Studying trend of email over time by mid

```
# Calculate the number of unique `mid` per date
trend_message <- CompleteDataset %>%
  group_by(date) %>%
  summarize(unique_mids = n_distinct(mid), .groups = 'drop')%>% # Count unique `mid`
  arrange(desc(unique_mids))

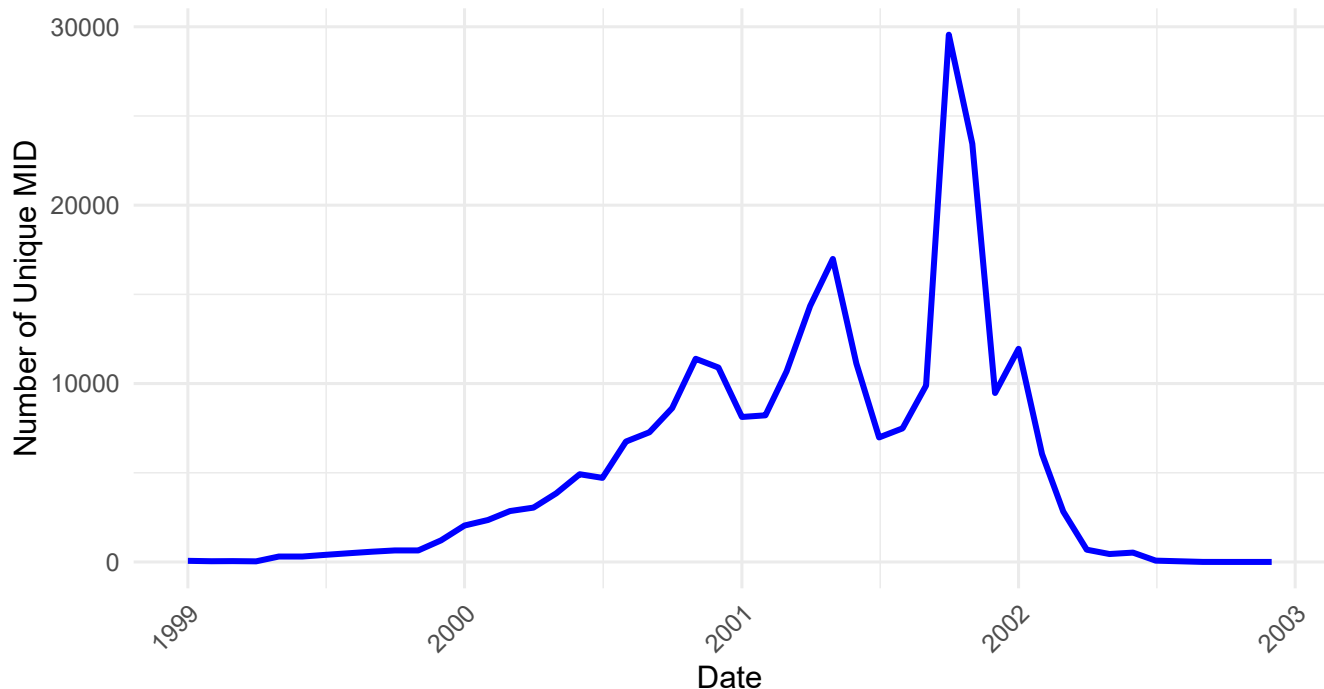
# Calculate the number of unique `mid` per month-year
trend_message_byMD <- trend_message %>%
  mutate(month_year = floor_date(date, "month")) %>%
  group_by(month_year) %>% # Group by month_year
  summarize(total_unique_mids = sum(unique_mids), .groups = 'drop') %>%
  arrange(desc(total_unique_mids))
trend_message_byMD
```

```
## # A tibble: 46 × 2
##   month_year total_unique_mids
##   <date>         <int>
## 1 2001-10-01      29556
## 2 2001-11-01      23441
## 3 2001-05-01      16986
## 4 2001-04-01      14348
## 5 2002-01-01      11948
## 6 2000-11-01      11388
## 7 2001-06-01      11128
## 8 2000-12-01      10899
## 9 2001-03-01      10686
## 10 2001-09-01       9888
## # i 36 more rows
```

```
# Plot the trend of unique message sent per date
trend_message_plot <- ggplot(data = trend_message_byMD, aes(x = month_year, y = total_unique_mids)) +
  geom_line(color = "blue", linewidth = 1) +
  labs(
    title = "Trend of Messent Sent Over Time",
    x = "Date",
    y = "Number of Unique MID"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1) # Rotate x-axis labels
  ,
    plot.margin = margin(t = 20, r = 20, b = 80, l = 20) # Increase bottom margin
  )

girafe(code=print(trend_message_plot))
```

## Trend of Messent Sent Over Time



### COMMENT :

Based on two previous plots, the top periods when the most email were sent, are the following **periods of interest**:

**Oct-Nov 2001: around 54 000 msg** - ( coincide with the period of Enron announcing to investor that it was going to post 638 billions of lost and discloses a 1.2bn dollars reduction in shareholder equity - the emergence of the accounting fraud- Enron stock price going from 90 dollars to 12dollars per share )

**March-April 2001: around 30 000 msg** - a number of analysts began to dig into the details of Enron's publicly released financial statements. Beginning of realizing the possible fraud

**January 2002 around 12 000 msg** - : Andersen admits employees disposed of Enron documents. FBI investigation. 1 month after Enron filling for bankruptcy protection in Dec 2001.

**Studying trend of email over time by Email adress use to send the email**



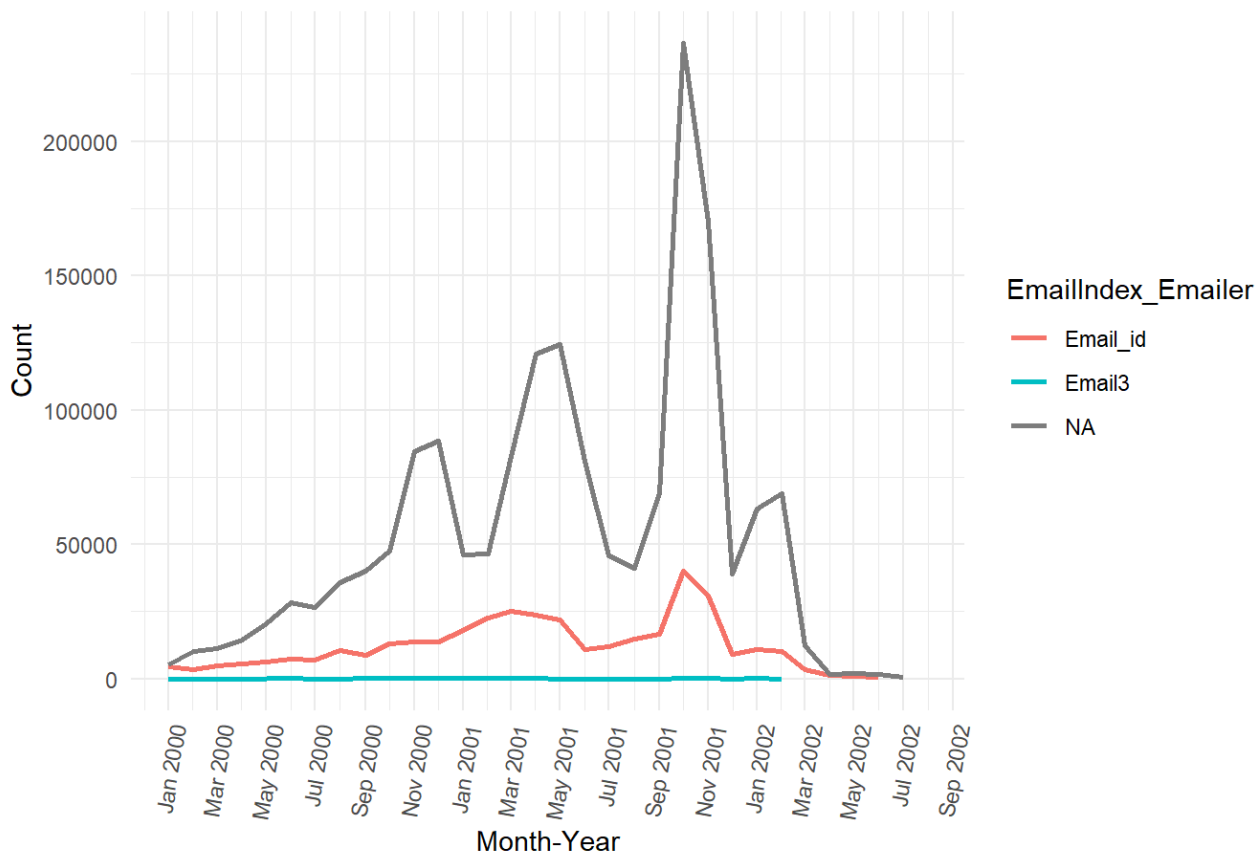
```

# Create a dataset summarizing by month_year and EmailIndex_Emailer
EmailIndex_Emailer_trend <- CompleteDataset %>%
  group_by(month_year, EmailIndex_Emailer) %>%
  summarize(count = n(), .groups = 'drop')

# Plot using the summarized data
ggplot(data = EmailIndex_Emailer_trend, aes(x = month_year, y = count, color = EmailIndex_Emailer, group = EmailIndex_Emailer)) +
  geom_line(size = 1) + # Use lines to show trends
  scale_x_date(
    date_labels = "%b %Y", # Format x-axis Labels as "Month Year"
    date_breaks = "2 month", # Break x-axis by every month
    limits = c(as.Date("2000-01-01"), as.Date("2002-07-31")) # Set date Limits
  ) +
  labs(
    title = "Counts Over Time by EmailIndex_Emailer",
    x = "Month-Year",
    y = "Count",
    color = "EmailIndex_Emailer"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 80, hjust = 1) # Rotate x-axis Labels for readability
  )

```

Counts Over Time by EmailIndex\_Emailer



COMMENT :

Even though the emails sent from all EmailIndex tend to follow the same trend, during periods of interest ,messages sent by non registered email addresses (meaning non-users) seems to have surge compared to others

We still see a slight difference around January - February where registered email stay steady when non-registered email increase a bit

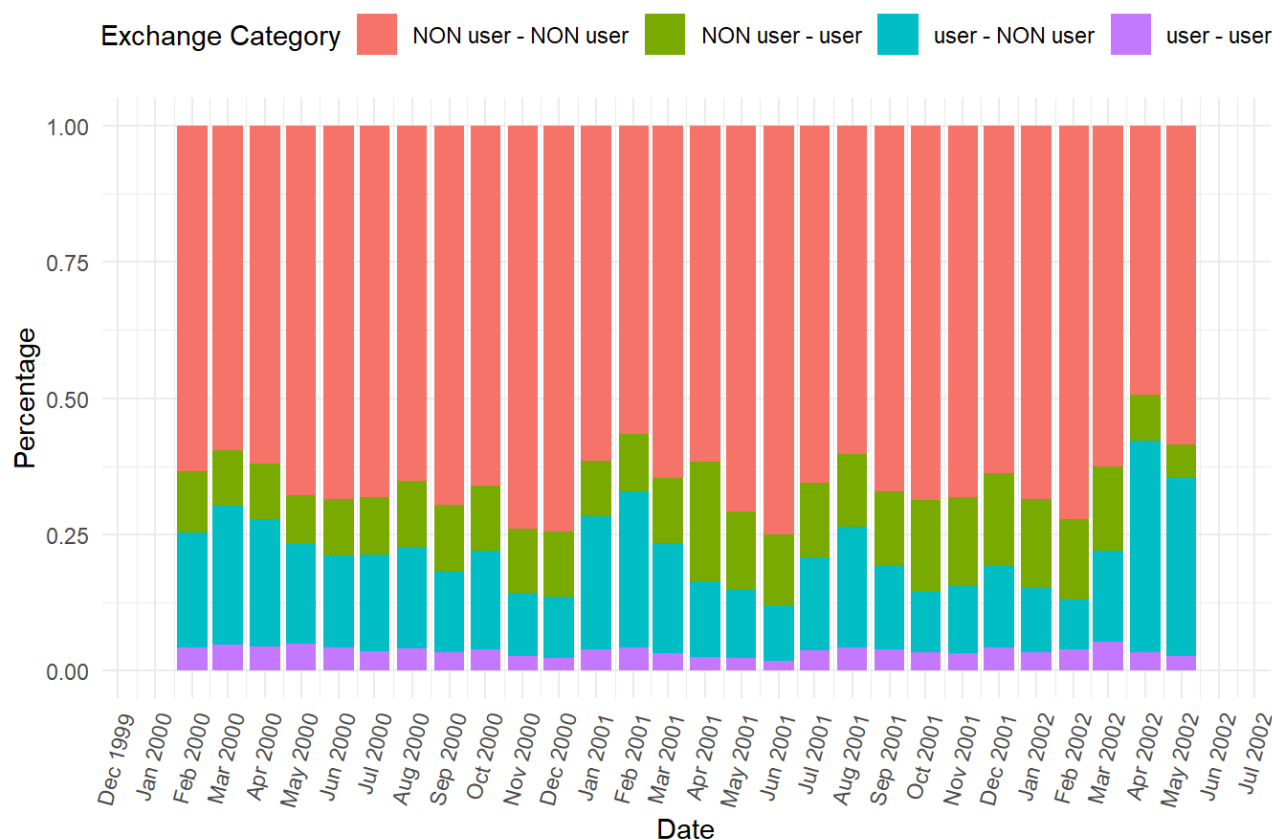
Note: this plot shows that user continue using their primary email address independently of the period.

### Studying trend of email over time between NON(user)/ (NON)user

```
ExchangeCat_trend <- CompleteDataset %>%
  group_by(month_year, ExchangeCat) %>%
  summarize(count = n(), .groups = 'drop') %>%
  mutate(percent = count / sum(count) * 100) # Calculate percentages

# Plot using the transformed data
ggplot(data = ExchangeCat_trend, aes(x = month_year, y = percent, fill = ExchangeCat)) +
  geom_bar(stat = "identity", position = "fill") + # 100% stacked bar chart
  scale_x_date(
    date_labels = "%b %Y", # Format x-axis labels as "Month Year"
    date_breaks = "1 month", # Break x-axis by every month
    limits = c(as.Date("2000-01-01"), as.Date("2002-05-31")) # Set date limits
  ) +
  labs(
    title = "100% Stacked Bar Chart of Exchange Categories Over Time",
    x = "Date",
    y = "Percentage",
    fill = "Exchange Category"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 75, hjust = 1), # Rotate x-axis labels for readability
    legend.position = "top" # Position legend at the top
  )
```

## 100% Stacked Bar Chart of Exchange Categories Over Time



### COMMENT:

The NON-user to NON-user exchange is consistently the most exchange category represented.

User to user exchanges seems to be stable

Even almost imperceptible, we can also notice that during the period of interest April 2001, messages coming from outside the registered user being send to user gained proportion

As far as user - non User exchanges are concerned, we can especially highlight their presence in April and May 2002. Possible event with link to "The DOJ indicts Arthur Andersen for obstruction of justice."

In April 2002 and May 2002, the proportion of emails sent by user outside of the organization has almost double vs the previous month

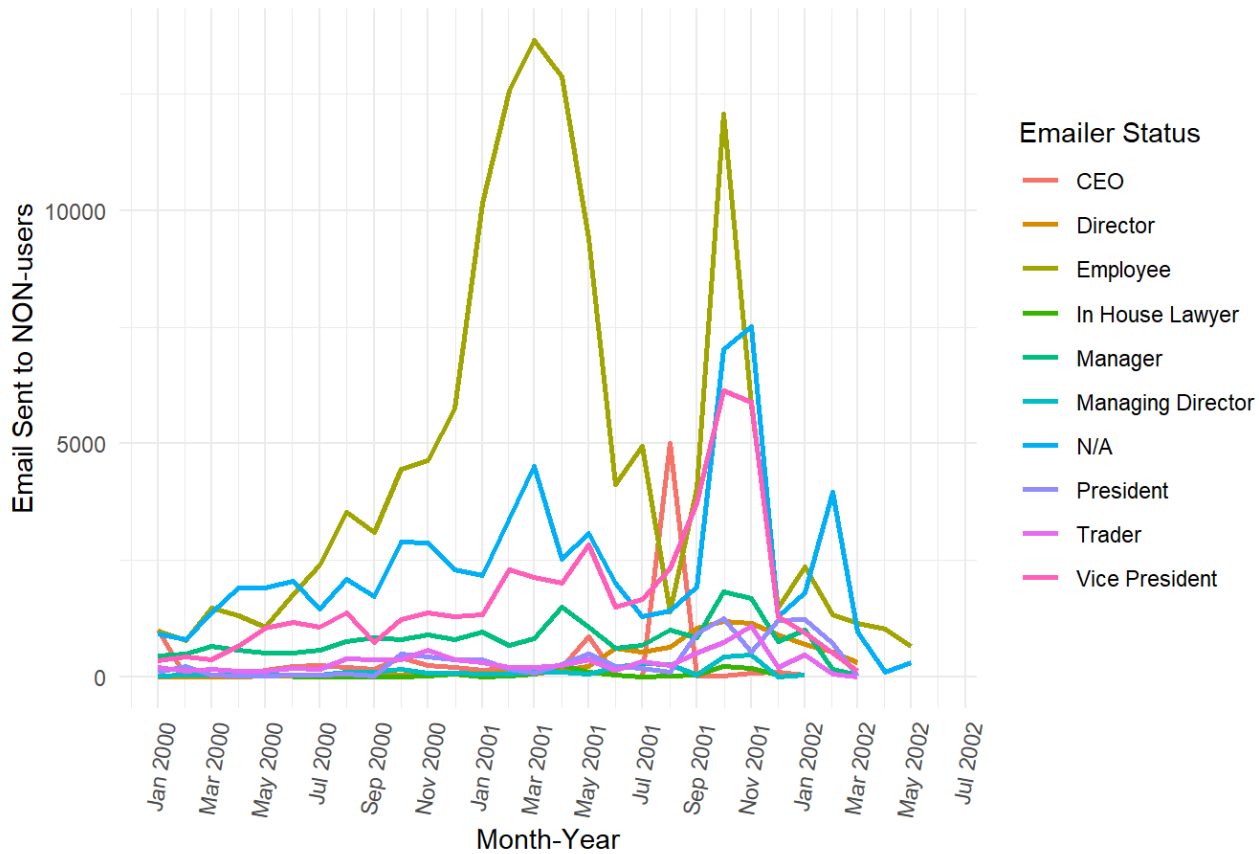
## Which status interacted the most with non registered user ?

```
StatusEmailingOut <- CompleteDataset %>%
  filter(ExchangeCat=="user - NON user" & EmailIndex_Emailer == "Email_id")%>%
  group_by (month_year,status_Emailer)%>%
  summarize(SentTo_NONUser= n(),.groups = "keep")
StatusEmailingOut
```

```
## # A tibble: 305 × 3
## # Groups:   month_year, status_Emailer [305]
##   month_year status_Emailer   SentTo_NONuser
##   <date>      <fct>          <int>
## 1 1999-01-01 Employee         111
## 2 1999-02-01 Employee          41
## 3 1999-03-01 Employee          53
## 4 1999-03-01 Vice President     1
## 5 1999-04-01 Employee          20
## 6 1999-05-01 Employee        100
## 7 1999-05-01 Managing Director    7
## 8 1999-05-01 N/A             304
## 9 1999-05-01 President           9
## 10 1999-05-01 Vice President     15
## # i 295 more rows
```

```
# Plot using the summarized data
ggplot(data = StatusEmailingOut, aes(x = month_year, y = SentTo_NONuser, color = status_Emailer, group = status_Emailer)) +
  geom_line(size = 1) + # Use lines to show trends
  scale_x_date(
    date_labels = "%b %Y", # Format x-axis labels as "Month Year"
    date_breaks = "2 month", # Break x-axis by every month
    limits = c(as.Date("2000-01-01"), as.Date("2002-05-31")) # Set date limits
  ) +
  labs(
    title = "Email sent to non registered users Over Time by status_Emailer",
    x = "Month-Year",
    y = "Email Sent to NON-users",
    color = "Emailer Status"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 80, hjust = 1) # Rotate x-axis labels for readability
  )
```

## Email sent to non registered users Over Time by status\_Emailer



**COMMENT :** The most interaction with non user was by Employee's and N/A's. However, all status stopped interacting with non user after March 2002. Only Employees and N/A's seem to have continued their interaction with people outside of the organisation until April 2002.

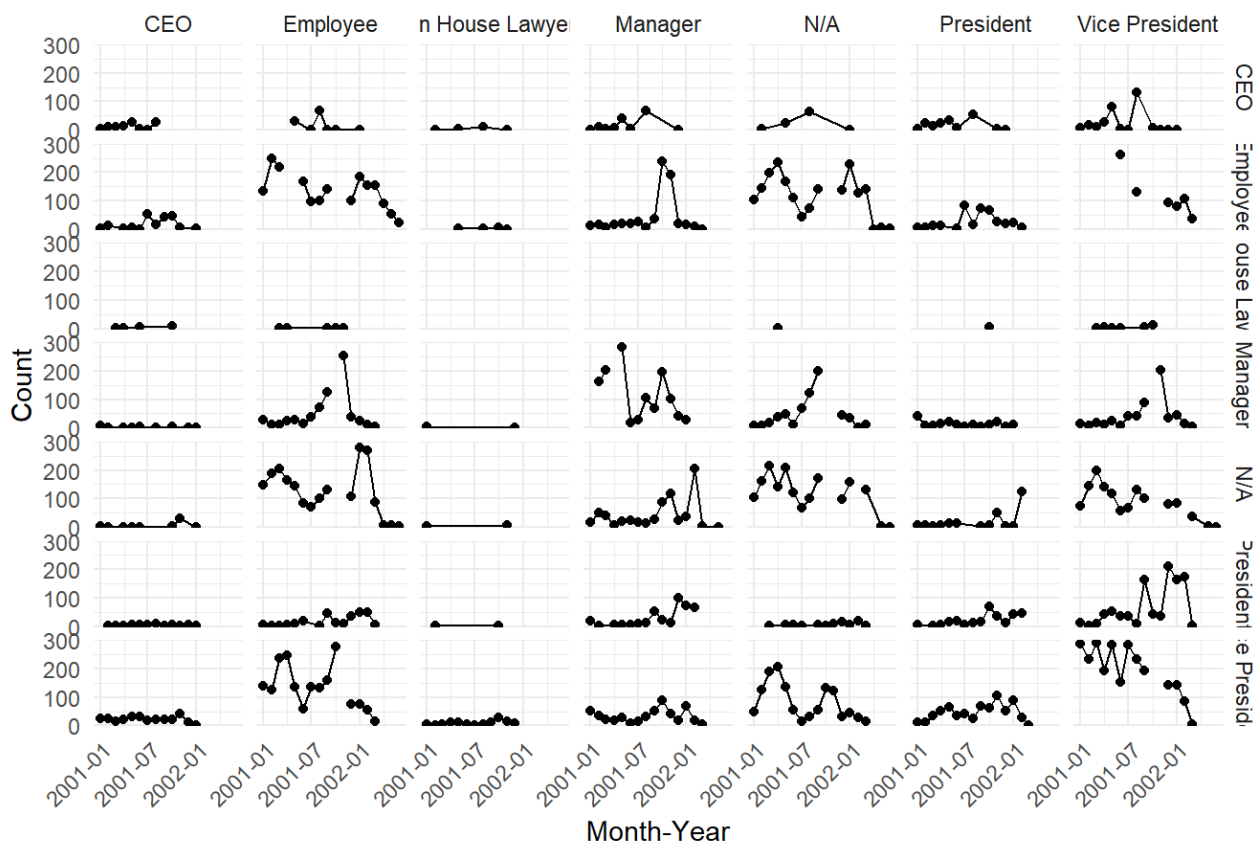
```

StatusVSStatusData <- CompleteDataset %>%
  # Convert month_year to Date if it is not already
  mutate(month_year = as.Date(month_year, format = "%Y-%m-%d")) %>% # Adjust the format as needed
  group_by(month_year, status_Emailer, status_Recipient) %>%
  filter(
    status_Emailer %in% c("President", "CEO", "Vice President", "In House Lawyer", "N/A", "Manager", "Employee") &
    status_Recipient %in% c("President", "CEO", "Vice President", "In House Lawyer", "N/A", "Manager", "Employee") &
    month_year >= as.Date("2001-01-01") &
    month_year <= as.Date("2002-06-30")
  ) %>%
  summarise(count = n(), .groups = 'keep')

ggplot(data=StatusVSStatusData)+
  geom_point(aes(x = month_year, y=count))+
  #geom_smooth(aes(x = month_year, y=count),method="lm")+
  geom_line(aes(x = month_year, y=count))+
  facet_grid(status_Emailer~status_Recipient)+
  scale_y_continuous(
    limits = c(0, 300), # Set the y-axis limits from 0 to 300
    expand = c(0, 0) # Prevent additional space beyond the limits
  ) +
  labs(
    title = "Interaction From Status_Emailer To Status_Recipient Over Time",
    x = "Month-Year", # Label for x-axis
    y = "Count" # Label for y-axis
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"), # Center and bold the title
    axis.text.x = element_text(angle = 45, hjust = 1) # Rotate x-axis labels by 45 degrees
  )

```

## Interaction From Status\_Emailer To Status\_Recipient Over Time



### COMMENT:

Possible interaction of interest:

-Vice President to CEO in August 2001

-Vice President to President in Sept 2001

- N/A to Manager between June 2001 and Sept 2001

-Manager to Employee between October 2001 and November 2001 They indeed show a change of dynamic during period previously flagged as “period of interest”

## OBJECTIVE 4: ANALYZE EMAILS CONTENT

### QUESTIONS RELATED TO OBJECTIVE 4

-Does the exchanges reflects that people were aware about the situation?

-Were there any information shared about possible layoffs?

-Content of Messages received or sent by Enron case person of interest if existent: Ken Lay (CEO), Jeff Skilling, Andrew Faston, Andersen Arthur

-Content of message of “possible interaction of interest”

Adding information to the previous general dataset to help us answer our interrogation for objective 4 part.

```

WordOfInterest <- "fraud|bankruptcy|layoffs|debt|lawsuit|litigation|resignation" # _Inserting to be re
search_

ContentCompleteDataset <- CompleteDataset %>%
  filter(!is.na(reference)) %>%
  mutate(
    word_exist = str_detect(tolower(reference), WordOfInterest), #binary column detecting the existen
ce of the word
    word_count = str_count(tolower(reference), WordOfInterest), # counting how many times the terms a
ppears
    word_extract = str_extract(tolower(reference), WordOfInterest) #extracting the first word found
  )

ContentCompleteDataset_TRUE <- ContentCompleteDataset %>%
  filter(word_exist == TRUE)

```

## Estimating a time range for these topics

```

#Estimating a time range for these topics
min(ContentCompleteDataset_TRUE$date)

```

```
## [1] "1999-09-23"
```

```
min(ContentCompleteDataset_TRUE$date[ContentCompleteDataset_TRUE$word_count>1])
```

```
## [1] "2000-05-22"
```

```
max(ContentCompleteDataset_TRUE$date)
```

```
## [1] "2002-06-18"
```

**COMMENT:** The topic first appear in the exchanges in Sept 1999. But it became more of a hot topic after May 2000. The topics were still mentioned in exchanges in June 2002.

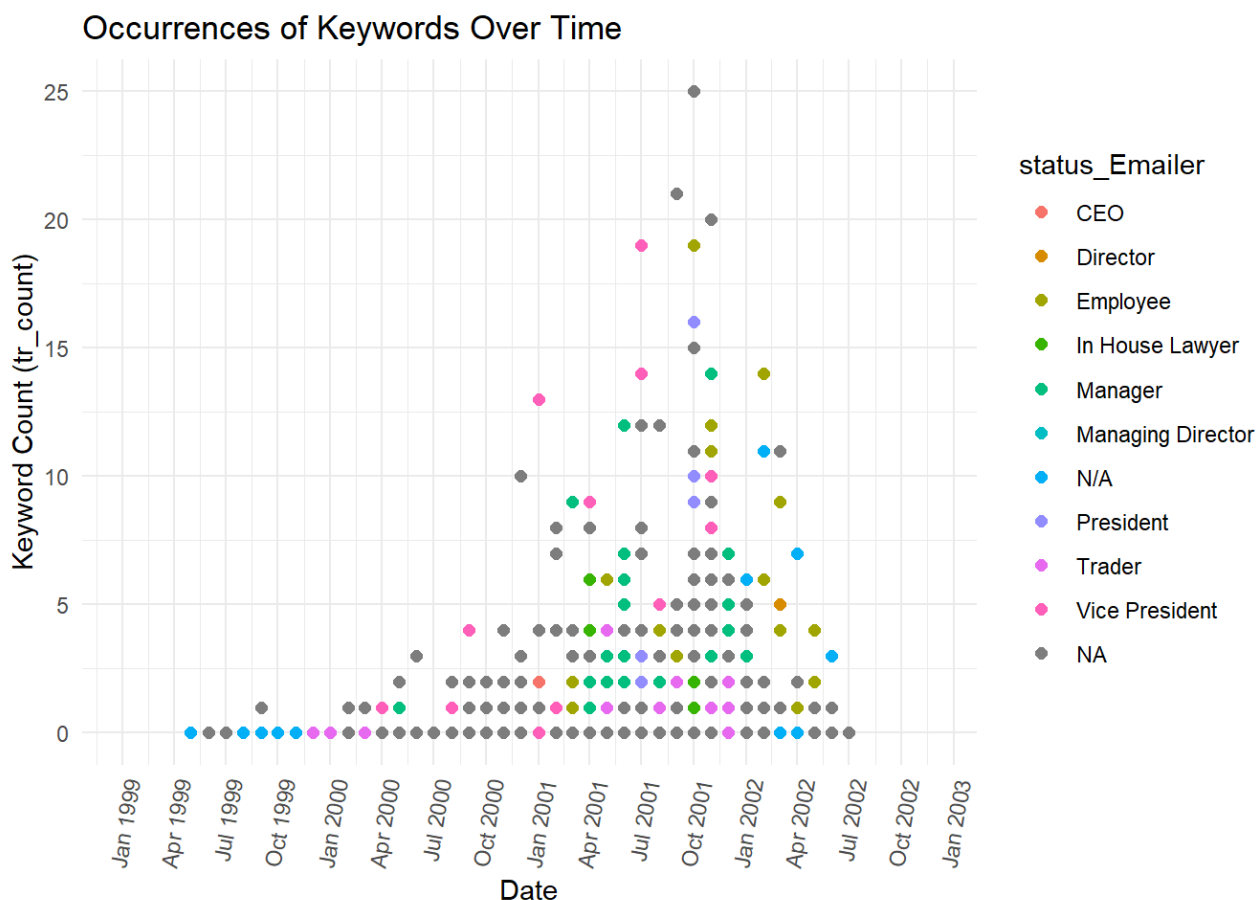


```

ContentMessage <- CompleteDataset %>% distinct(mid, month_year, subject, reference, fullName_Emailer, status_Emailer, fullName_Recipient, status_Recipient) %>% mutate(word_count = str_count(tolower(reference), WordOfInterest))

ggplot(ContentMessage, aes(x = month_year, y = word_count, color = status_Emailer)) +
  geom_point(size = 2) +
  scale_x_date(
    date_labels = "%b %Y", # Format x-axis labels as "Month Year"
    date_breaks = "3 month", # Break x-axis by every month
  ) +
  labs(
    title = "Occurrences of Keywords Over Time",
    x = "Date",
    y = "Keyword Count (tr_count)"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 80, hjust = 1) # Rotate x-axis labels for readability
  )

```



**COMMENT:** As expected the key words follows the trend of the period of interest and tension at Enron

## Determining the status and exchanges were the company's situation

## was the most discuss.

```
ContentCompleteDataset_TRUE %>%
  filter(word_count>=1) %>%
  group_by(ExchangeStatus)%>%
  summarize(NbExchange=n())%>%
  arrange(desc(NbExchange))
```

```
## # A tibble: 80 × 2
##   ExchangeStatus      NbExchange
##   <chr>              <int>
## 1 NA-NA              3773
## 2 Employee-NA       1248
## 3 Vice President-NA  549
## 4 NA-Vice President  372
## 5 NA-Employee        337
## 6 N/A-NA             314
## 7 NA-N/A             163
## 8 Manager-NA         120
## 9 Employee-Vice President  99
## 10 NA-President      93
## # i 70 more rows
```

**COMMENT:** The topics were mostly at heart of the exchanges involving three status whether they were emailers or recipients :

- NAs
- Employees
- Vice Presidents

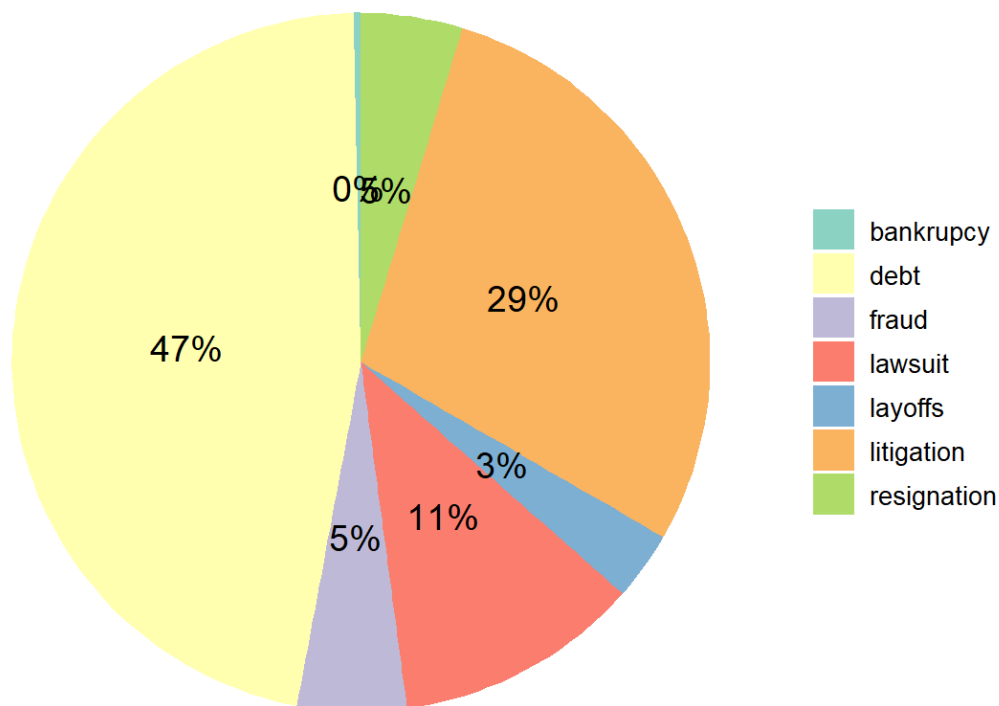
## Distribution of the words of interest in the conversations

```
WordPieData <- ContentCompleteDataset_TRUE %>% group_by(word_extract) %>% summarise(WordExtractedCount = n()) %>% arrange(desc(WordExtractedCount))%>% mutate(WordExtractedCountPercentage= round(WordExtractedCount/sum(WordExtractedCount)*100))
WordPieData
```

```
## # A tibble: 7 × 3
##   word_extract WordExtractedCount WordExtractedCountPercentage
##   <chr>          <int>              <dbl>
## 1 debt          3756              47
## 2 litigation    2301              29
## 3 lawsuit       915              11
## 4 fraud         405               5
## 5 resignation   378               5
## 6 layoffs       255               3
## 7 bankruptcy    26                0
```

```
# Pie representation of the distribution**
ggplot(data = WordPieData, aes(x = '', y = WordExtractedCount, fill = word_extract)) +
  geom_bar(stat = 'identity', width = 1) + # Create bars to form a pie
  geom_text(aes(label = paste0(WordExtractedCountPercentage, '%')),
            position = position_stack(vjust = 0.5), color = 'black', size = 5) + # Add percentage Labels
  coord_polar(theta = 'y') + # Turn the bar chart into a pie chart
  theme_void() + # Remove background
  scale_fill_brewer(palette = "Set3") + # Choose a color palette for better readability
  theme(legend.title = element_blank(), # Remove Legend title for clarity
        legend.text = element_text(size = 10)) + # Adjust Legend text size
  ggtitle("Distribution of Extracted Words")
```

Distribution of Extracted Words



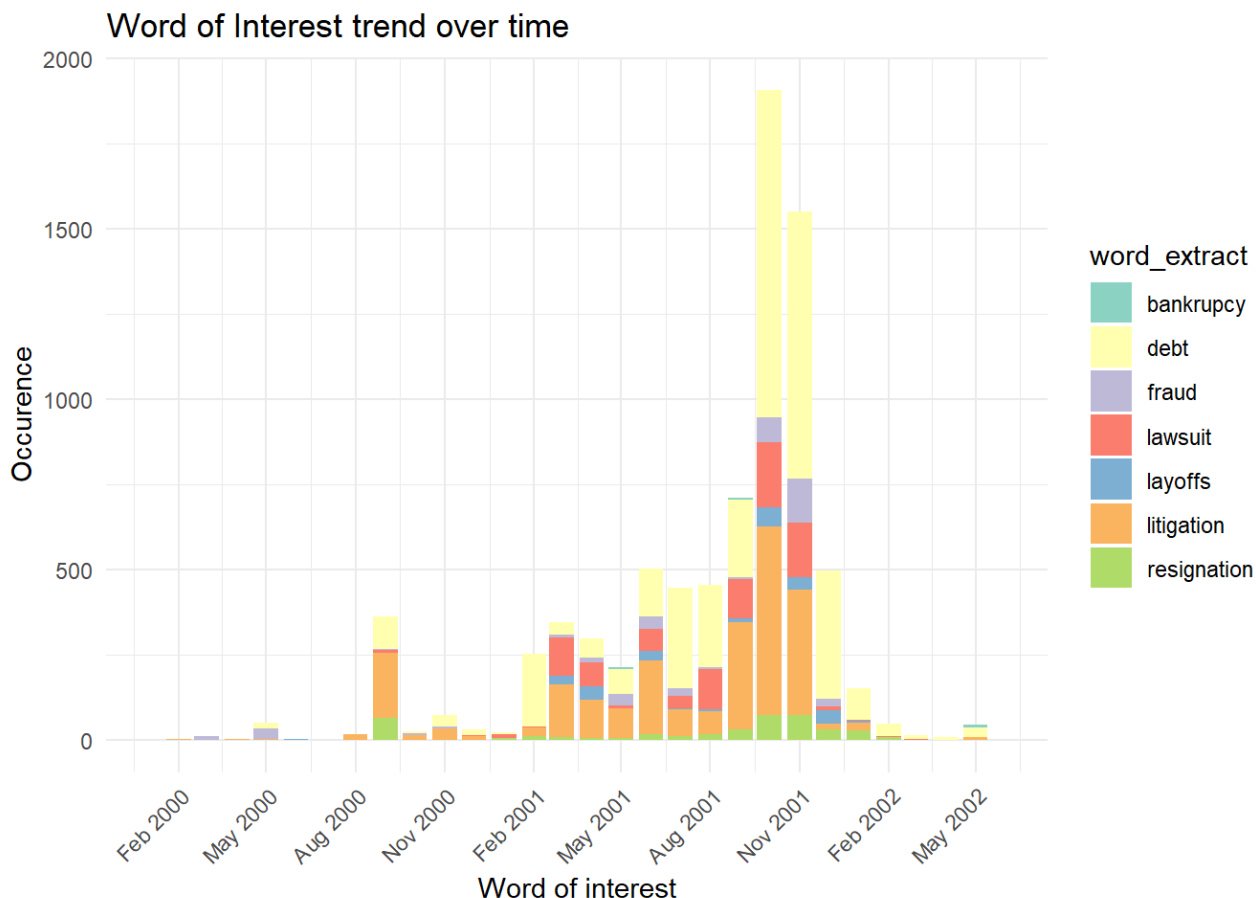
**COMMENT :** Most conversation are related to debt and litigation

```

WordLineTrendData <- ContentCompleteDataset_TRUE %>%
  select(month_year, word_extract) %>%
  group_by(month_year, word_extract) %>%
  mutate(WordExtractedCount = n()) %>%
  arrange(month_year) %>%
  distinct(month_year, word_extract, .keep_all = TRUE)

ggplot(WordLineTrendData, aes(x = month_year, y = WordExtractedCount, fill = word_extract)) +
  geom_bar(stat = "identity") +
  scale_x_date(
    date_labels = "%b %Y", # Format x-axis Labels as "Month Year"
    date_breaks = "3 month", # Break x-axis by every month
    limits = c(as.Date("2000-01-01"), as.Date("2002-05-31")) # Set date Limits
  ) +
  labs(title = "Word of Interest trend over time", x = "Word of interest", y = "Occurence") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set3")+ theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



**COMMENT:** During October 2001 and November 2001, period the most tense for the company, the topic off debts, fraud and lawsuit became predominant. We also observe a rise in resignations and layoff.

*As a recall, in Oct-Nov 2001 : " Enron announcing to investor that it was going to post 638 billions of lost and discloses a 1.2bn dollars reduction in shareholder equity - the emergence of the accounting fraud- Enron stock price going from 90 dollars to 12dollars per share"*

The topic of fraud start becoming visible in May-July 2001.

*As a recall, " in march - April 2001, number of analysts began to dig into the details of Enron's publicly released financial statements. Beginning of realizing the possible fraud"*