

Word Ranking of the Holy Quran in the base of Principal Component Analysis in R.

Sk Mijanur Rahaman

16/07/2021

#Question: What are the most frequent and most dispersed as well as most frequent but less dispersed words,bigrams and trigrams in the Holy Quran?

#To find out the rank of words,bigrams and trigrams in the Holy Quran data set in R based on Principal component analysis i.e which words, bigrams and trigrams have most as well as less frequency and which have most as well as less variability in the Quran for both arabic and english languages.

#Procedure: As the data set available in the library "quRan". Now we create new 6 data sets.Each data sets contain frequencies of each words (similar for bigrams and trigrams) in every chapters.

Therefore the new each data set's each columns contain chapters names and each rows contain the words(similar for bigrams and trigrams) for both arabic and english languages.

#Then we perform Principal Component Analysis in the multivariate data set

#of the frequency distribution of the each ngrams.

#As we know that the first Principal Component explain the maximum variability

#in the data. So we order the ngrams with respect to the first Principal

#Component. Then we would get the most frequent and most varied ngrams in the

#top of the order and most frequent but less varied ngrams in the bottom

#of the order.

```
library(quRan)
```

```
library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 4.1.2
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr 0.3.4
```

```
## v tibble 3.1.6       v dplyr 1.0.7
```

```
## v tidyr 1.1.4        v stringr 1.4.0
```

```
## v readr 2.1.0        v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.1.2
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

```
library(arabicStemR)
```

```
##
## Attaching package: 'arabicStemR'
```

```
## The following object is masked from 'package:graphics':
##
## stem
```

```
Sys.setlocale("LC_CTYPE","arabic")
```

```
## [1] "Arabic_Saudi Arabia.1256"
```

```
#To ranke of the Holy quran words in arabic:
#####
book_id=rep(1,nrow(quran_ar_min))
quran_ar_min1=cbind(book_id,quran_ar_min)
head(quran_ar_min1)
```

```
## book_id surah_id ayah_id surah_title_ar surah_title_en surah_title_en_trans
## 1 1 1 1 الفاتحة Al-Faatiha The Opening
## 2 1 1 2 الفاتحة Al-Faatiha The Opening
## 3 1 1 3 الفاتحة Al-Faatiha The Opening
## 4 1 1 4 الفاتحة Al-Faatiha The Opening
## 5 1 1 5 الفاتحة Al-Faatiha The Opening
## 6 1 1 6 الفاتحة Al-Faatiha The Opening
## revelation_type text surah ayah ayah_title juz
## 1 Meccan <U+FEFF>1 1:1 1 1 بسم الله الرحمن الرحيم
## 2 Meccan 1 1:2 2 1 الحمد لله رب العالمين
## 3 Meccan 1 1:3 3 1 الرحمن الرحيم
## 4 Meccan 1 1:4 4 1 مالك يوم الدين
## 5 Meccan 1 1:5 5 1 إياك نعبد وإياك نستعين
## 6 Meccan 1 1:6 6 1 اهدنا الصراط المستقيم
## manzil page hizb_quarter sajda sajda_id sajda_recommended sajda_obligatory
## 1 1 1 1 FALSE NA NA NA
## 2 1 1 1 FALSE NA NA NA
## 3 1 1 1 FALSE NA NA NA
## 4 1 1 1 FALSE NA NA NA
## 5 1 1 1 FALSE NA NA NA
## 6 1 1 1 FALSE NA NA NA
```

```
quran_words <- quran_ar_min1 %>%
  unnest_tokens(word, text) %>%
  count(book_id, word, sort = TRUE)

colnames(quran_words)=c("book_id","word","freq_in_book")

head(quran_words)
```

```
## book_id word freq_in_book
```

## 1	1	2763	من
## 2	1	2265	الله
## 3	1	1185	في
## 4	1	1010	ما
## 5	1	966	إن
## 6	1	812	لا

```
chapter_words <- quran_ar_min1 %>%
  unnest_tokens(word, text) %>%
  count(surah_title_en, word, sort = TRUE)

colnames(chapter_words)=c("surah_title_en","word","freq_in_chapter")

chapter_words=na.omit(chapter_words)
head(chapter_words)
```

```
## surah_title_en word freq_in_chapter
```

## 1	Al-Baqara	219	من
## 2	Al-Baqara	217	الله
## 3	An-Nisaa	184	الله
## 4	Aal-i-Imraan	154	الله
## 5	Aal-i-Imraan	145	من
## 6	Al-A'raaf	128	من

```
quran_words=left_join(quran_words,chapter_words)
```

```
## Joining, by = "word"
```

```
head(quran_words)
```

```
## book_id word freq_in_book surah_title_en freq_in_chapter
```

## 1	1	2763	من	Al-Baqara	219
## 2	1	2763	من	Aal-i-Imraan	145
## 3	1	2763	من	Al-A'raaf	128
## 4	1	2763	من	An-Nisaa	122
## 5	1	2763	من	Al-Maaida	93
## 6	1	2763	من	Al-An'aam	90

```
token_list=as.data.frame(table(chapter_words$surah_title_en,chapter_words$word))
```

```
token_list=token_list[,-3]
colnames(token_list)=c("surah_title_en","word")
head(token_list)

## surah_title_en word
## 1 Al-Faatiha ن
## 2 Al-Baqara ن
## 3 Aal-i-Imraan ن
## 4 An-Nisaa ن
## 5 Al-Maaida ن
## 6 Al-An'aam ن

head(quran_words)

## book_id word freq_in_book surah_title_en freq_in_chapter
## 1 1 2763 من Al-Baqara 219
## 2 1 2763 من Aal-i-Imraan 145
## 3 1 2763 من Al-A'raaf 128
## 4 1 2763 من An-Nisaa 122
## 5 1 2763 من Al-Maaida 93
## 6 1 2763 من Al-An'aam 90

quran_words=left_join(token_list,quran_words,by=c("surah_title_en","word"))

quran_words[is.na(quran_words)]=0
head(quran_words)

## surah_title_en word book_id freq_in_book freq_in_chapter
## 1 Al-Faatiha 0 0 0 ن
## 2 Al-Baqara 0 0 0 ن
## 3 Aal-i-Imraan 0 0 0 ن
## 4 An-Nisaa 0 0 0 ن
## 5 Al-Maaida 0 0 0 ن
## 6 Al-An'aam 0 0 0 ن

arabic_stopwords <- data.frame(word = removeStopWords("سلام")$arabicStopwordList)

quran_words=anti_join(quran_words,arabic_stopwords, by = "word")

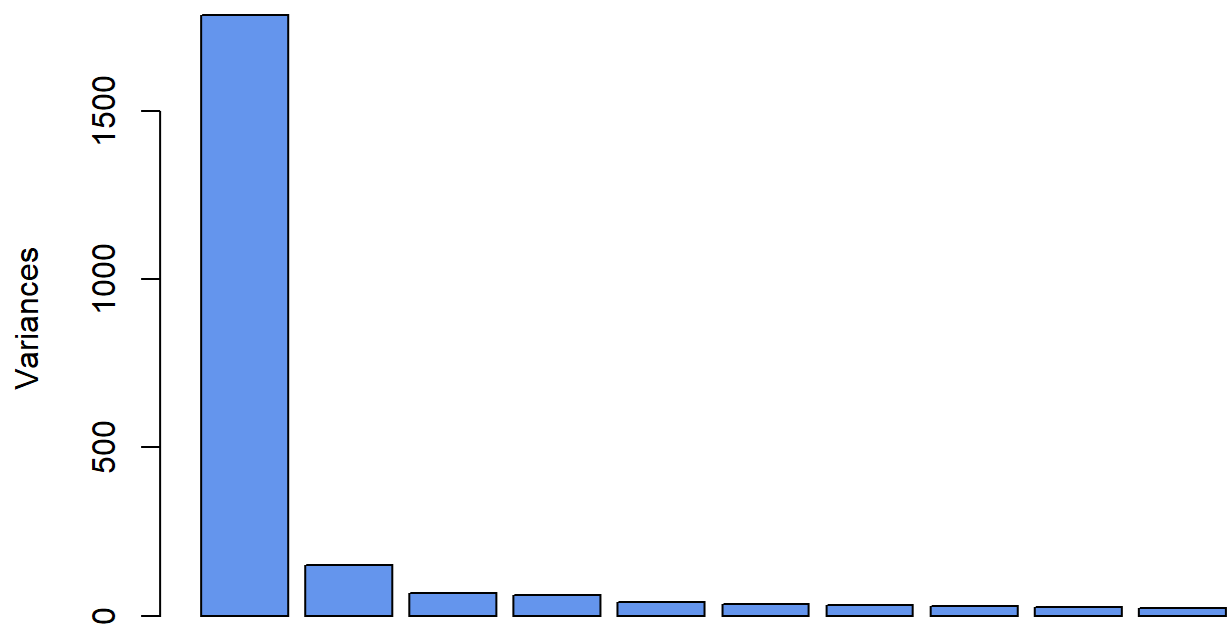
r=unique(quran_words$word)
c=unique(quran_ar_min$surah_id)

#Frequency distribution of each words per each chapters:
dword=matrix(quran_words$freq_in_chapter,nrow =nrow(quran_words)/114 ,
            ncol = 114,byrow = T,dimnames = list(r,c))

head(dword)
```

```
screepLOT(pca1,col="cornflowerblue",main = "Scree plot of pca for  
word distn in arabic.")
```

Scree plot of pca for word distn in arabic.



```
#To rank the we sort with respect to first principal components:
ranking=pca1[[2]][order(pca1[[2]][,1],decreasing = TRUE),]
head(ranking[,1],10)
```

```
## الله الذين والله آمنوا فإن كان قال
## 0.85109424 0.28232317 0.12017018 0.10223861 0.09346210 0.08589802 0.08317635
## قل الكتاب الناس
## 0.07909102 0.07734340 0.07004880
```

```
tail(ranking[,1],10)
```

```
## الإنسان للمكذبين أدراك كلا ويل الرحمن
## -0.001112508 -0.001319836 -0.001433864 -0.001451245 -0.001468456 -0.001952542
## آلاء ربكما فبأي تكذبان
## -0.003103474 -0.003342187 -0.003515027 -0.003557830
```

```
tail1=tail(ranking,20)
ranked=rownames(ranking)
```

```
#After ranking we got in top ranked in Arabic:
head(ranked,100)
```

```
## [1] "الله" "الذين" "ولله" "آمنوا" "فإن" "كان"
## [7] "قال" "قل" "الكتاب" "الناس" "قالوا" "بالله"
## [13] "كنتم" "الأرض" "أيها" "كفروا" "كل" "فلا"
## [19] "شيء" "كانوا" "أنزل" "أولئك" "والذين" "عليهم"
## [25] "خير" "سبيل" "الدنيا" "وهو" "الله" "يشاء"
## [31] "السموات" "عذاب" "النار" "الحق" "ربنا" "أنفسهم"
## [37] "إبراهيم" "وأنتم" "منه" "موسى" "وكان" "ربكم"
## [43] "فقد" "يوم" "لعلكم" "المؤمنين" "تعملون" "منها"
## [49] "واتقوا" "ربهم" "يؤمنون" "الصلاة" "أهل" "والأرض"
## [55] "يعلمون" "شيئا" "للناس" "رحيم" "ورسوله" "دون"
## [61] "قوم" "غفور" "قلوبهم" "تعملون" "يحب" "ربك"
## [67] "القوم" "الشيطان" "العذاب" "ألم" "الرسول" "غير"
## [73] "جناح" "إليك" "الكافرين" "أنفسكم" "القيامة" "الآخرة"
## [79] "إليه" "الآخر" "قليل" "فإذا" "للذين" "كثيرا"
## [85] "فأولئك" "مريم" "النساء" "بالمعروف" "بعض" "يعلم"
## [91] "الحرام" "يعملون" "جميعا" "وعلى" "الظالمين" "مثل"
## [97] "ولقد" "إله" "يهدي" "يريد"
```

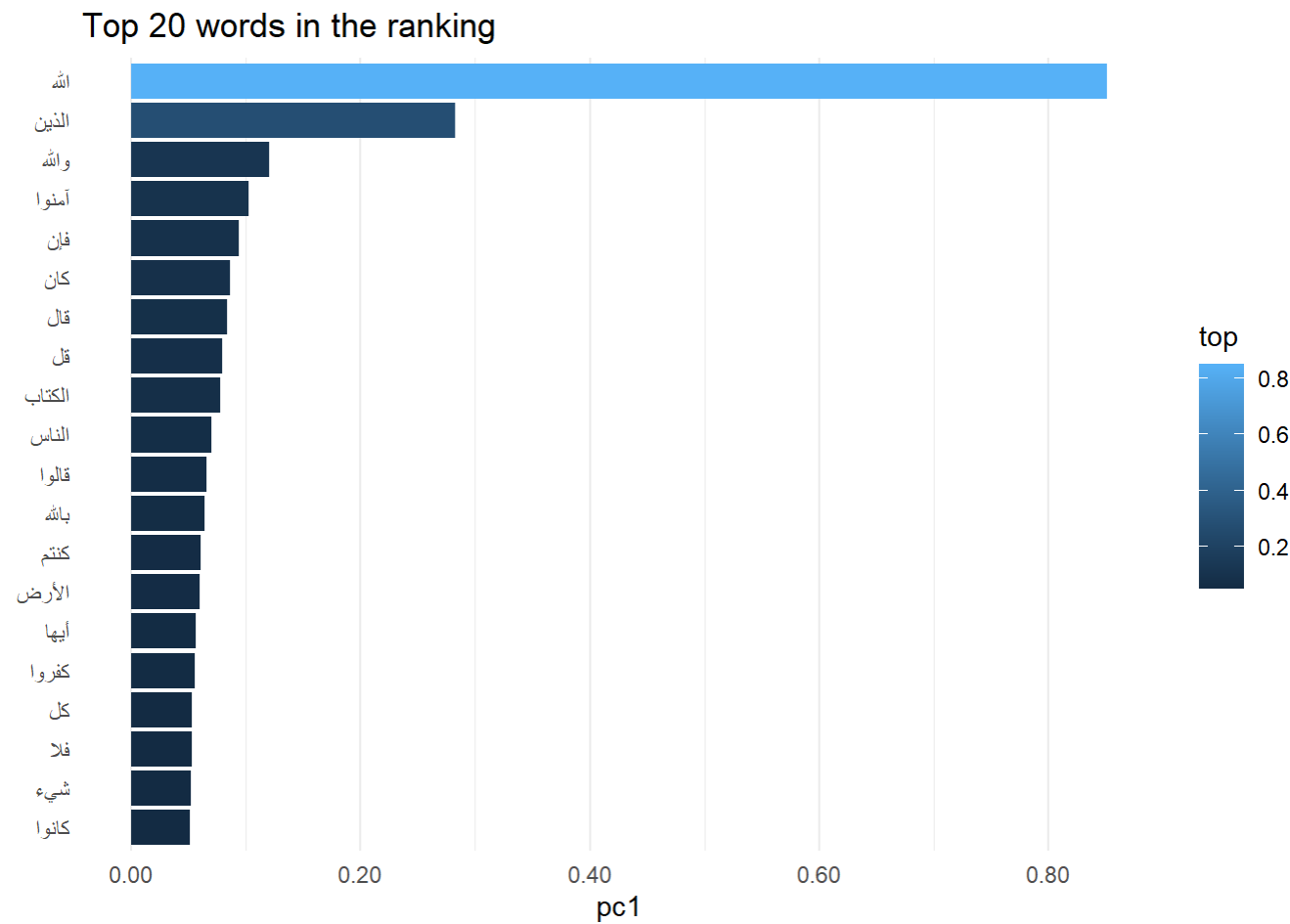
```
#After ranking we got in last ranked in Arabic:
tail(ranked,100)
```

```
## [1] "الحاقة" "يسرا" "القول" "يره" "فضة" "القدر"
## [7] "إنس" "الميزان" "ينبغي" "حسابا" "القارعة" "كيدا"
## [13] "الجبال" "رجيم" "سلاما" "ممنون" "وبنين" "صلصال"
## [19] "يتساءلون" "أفرأيتم" "نبيا" "ويلنا" "الأرائك" "المرسلون"
## [25] "سقر" "عذابي" "نادى" "أدبر" "آخر" "المأوى"
## [31] "بخشى" "الصراط" "ورب" "ترابا" "أواب" "يغنى"
## [37] "بالنذر" "الشمال" "فذكر" "مصبحين" "ونذر" "يكذب"
## [43] "جند" "بسم" "أنا" "المسكين" "يسرنا" "سرر"
## [49] "باسم" "وكننا" "فكانت" "أإذا" "راضية" "أحسب"
## [55] "وتواصوا" "خاشعة" "كتابه" "صفا" "للرحمن" "أرأيت"
## [61] "صيحة" "وعيون" "مجنون" "الأولى" "وكل" "نجعل"
## [67] "الفصل" "طغى" "أتاك" "يوعدون" "وتولى" "سبحان"
## [73] "تبصرون" "الكبرى" "فسبح" "قدر" "فاكهة" "متكئين"
## [79] "معلوم" "اليقين" "اليمين" "الآخرين" "مذكر" "عين"
## [85] "تذكرة" "والليل" "يومئذ" "تكذبون" "أقسم" "خلقنا"
## [91] "الإنسان" "للمكذبين" "أدراك" "كلا" "ويل" "الرحمن"
## [97] "آلاء" "ربكما" "فبأي" "تكذبان"
```

```
#To plot Top 20 words in the ranking
```

```
ranking1=data.frame(x1=rownames(ranking)[1:20], top=ranking[1:20,1])

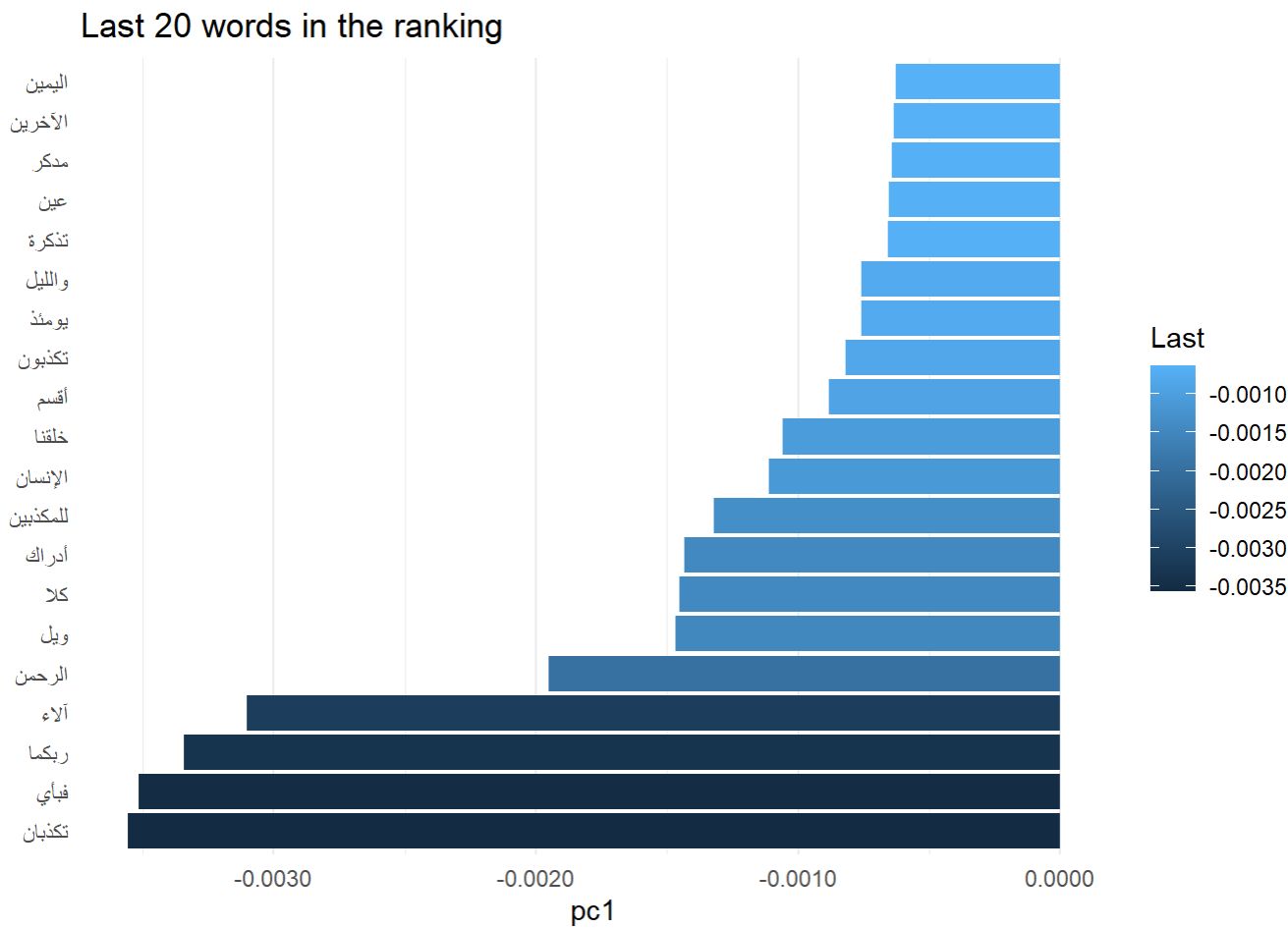
ggplot(ranking1, aes(x = fct_rev(reorder(x1,-top)),
                      y = top, fill = top)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Top 20 words in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
#To plot Last 20 words in the ranking

ranking2=data.frame(x1=rownames(tail1),Last=tail1[,1])

ggplot(ranking2, aes(x = fct_rev(reorder(x1,-Last)),
                    y = Last, fill = Last)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Last 20 words in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

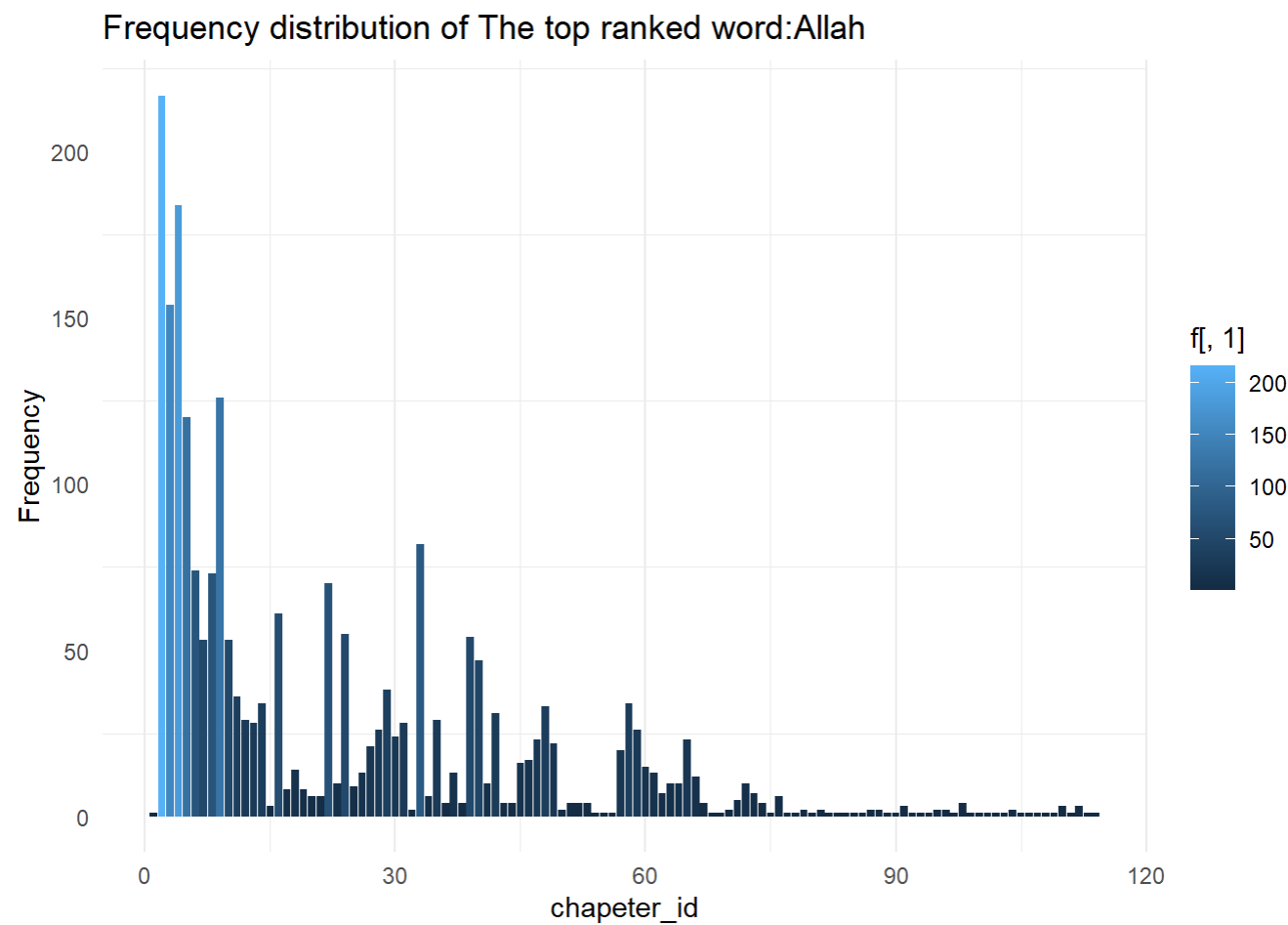
```
dword1=as.data.frame(dword)

f=as.matrix(filter(dword1 ,rownames(dword1) %in% head(ranked,1)))
fl=as.matrix(filter(dword1 ,rownames(dword1) %in% tail(ranked,1)))

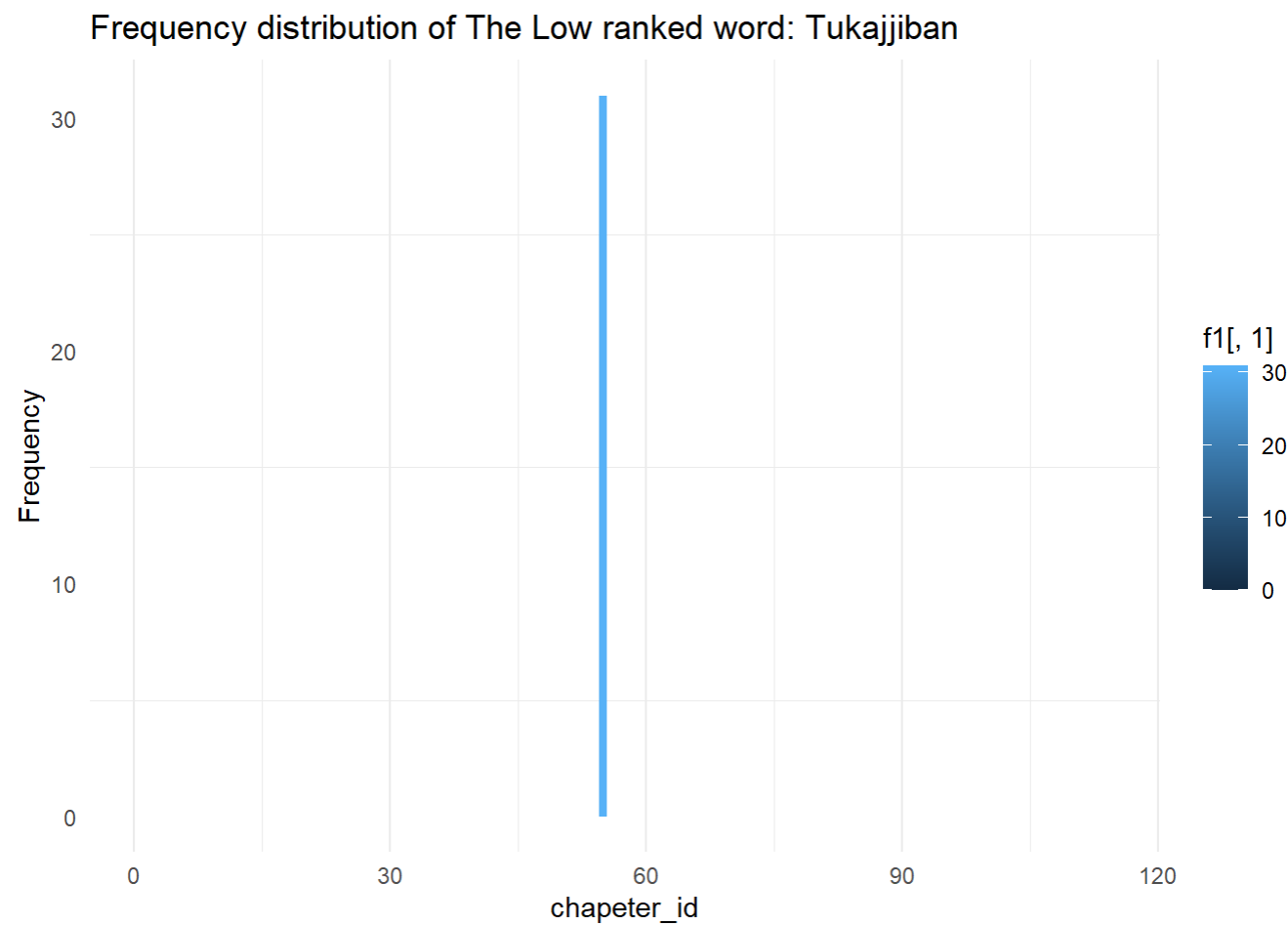
f=as.data.frame(t(f))
f=cbind(f,chapter_id=1:114)

fl=as.data.frame(t(fl))
fl=cbind(fl,chapter_id=1:114)

#Frequency distribution of The top ranked word
ggplot(f,aes(fill=f[,1],x=chapter_id,y=f[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The top ranked word:Allah") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
#Frequency distribution of The Low ranked word
ggplot(f1,aes(fill=f1[,1],x=chapter_id,y=f1[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapter_id",title = "Frequency distribution of The Low ranked word: Tukajjiban") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
#####  
#Manually ranking i.e univariately:  
  
dword11=data.frame(cbind(dword,rowSums(dword)))  
dword11$zeros=rowSums(dword11==0)  
dword11$occur=114-(dword11$zeros)  
dword11$combine=(.4*dword11$V115)+(.6*dword11$occur)  
write.table(dword11, file = 'posts.csv',sep = "\t", fileEncoding = "UTF-8")  
#=dw  
write.table(dword11,row.names = T,file = "wordfreq1.csv",sep = "\t",fileEncoding = "UTF-8")  
  
ranked=rownames(dword11[order(dword11$combine, decreasing = TRUE),])  
  
head(ranked)
```

```
## [1] "الله" "الذين" "قال" "كان" "الأرض" "قل" [1]
```

```
tail(ranked)
```

```
## [1] "ففدية" "ففروا" "ففريقا" "ففرت" "ففسق" "ففسقوا" [1]
```

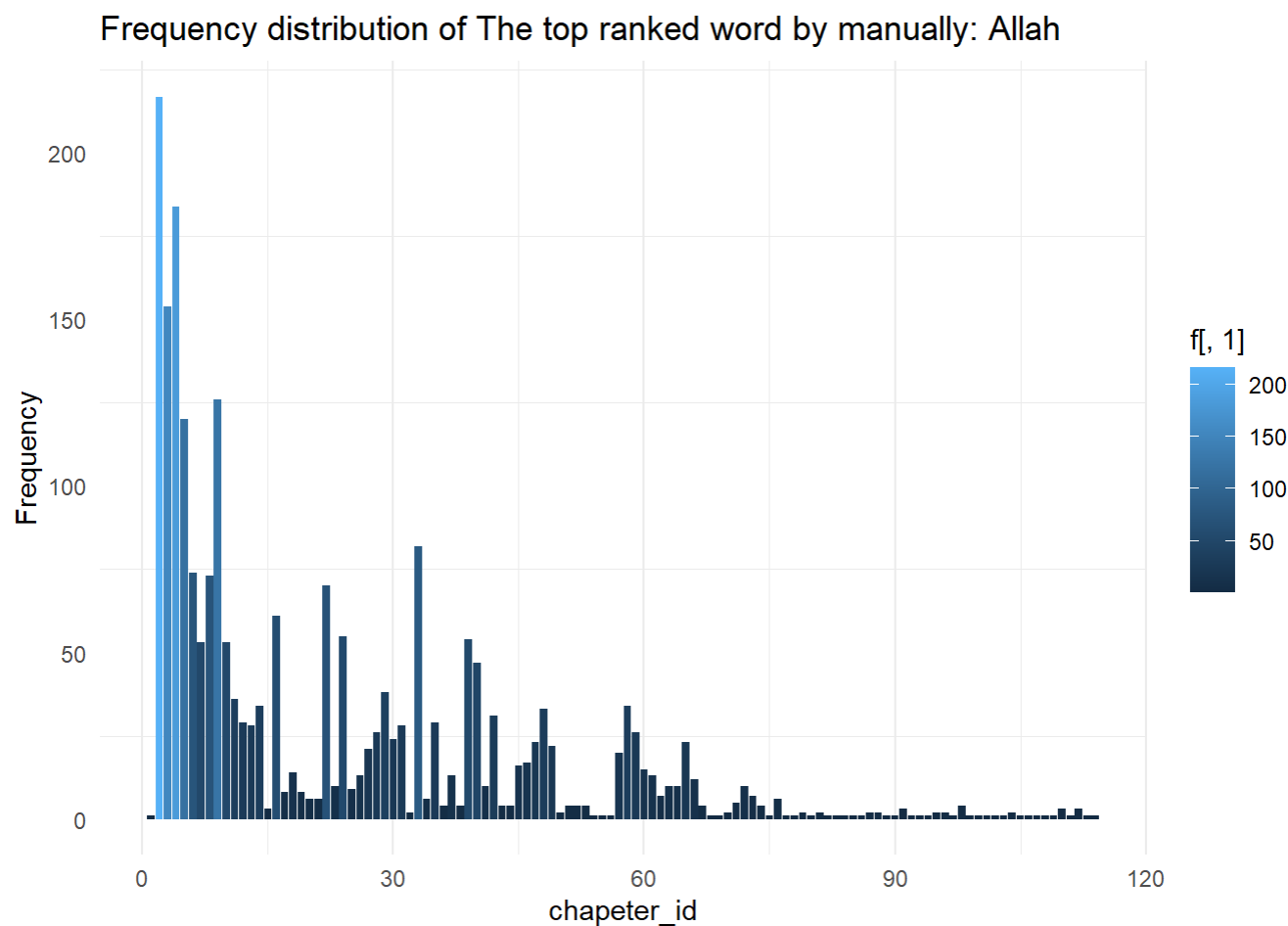
```
dword13=as.data.frame(dword11)
```

```
f=as.matrix(filter(dword1 ,rownames(dword1) %in% head(ranked,1)))
f1=as.matrix(filter(dword1 ,rownames(dword1) %in% tail(ranked,1)))

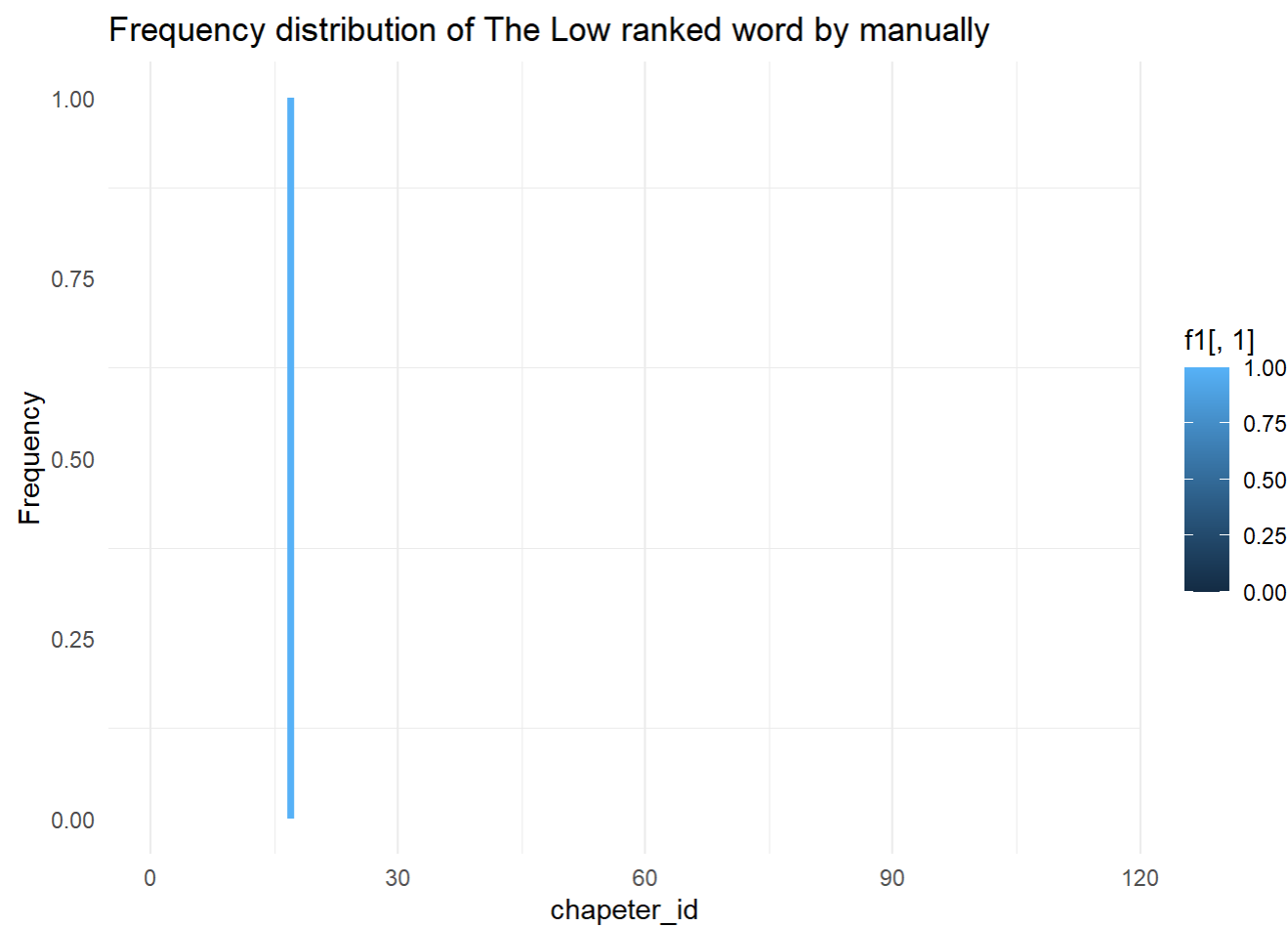
f=as.data.frame(t(f))
f=cbind(f,chapter_id=1:114)

f1=as.data.frame(t(f1))
f1=cbind(f1,chapter_id=1:114)

#Frequency distribution of The top ranked word
ggplot(f,aes(fill=f[,1],x=chapter_id,y=f[,1]))+
  geom_col()+
  labs(y = "Frequency", x = "chapeter_id",title = "Frequency distribution of The top ranked word by manually: Allah") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
#Frequency distribution of The Low ranked word
ggplot(f1,aes(fill=f1[,1],x=chapter_id,y=f1[,1]))+
  geom_col()+
  labs(y = "Frequency", x = "chapeter_id",title = "Frequency distribution of The Low ranked word by manually") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

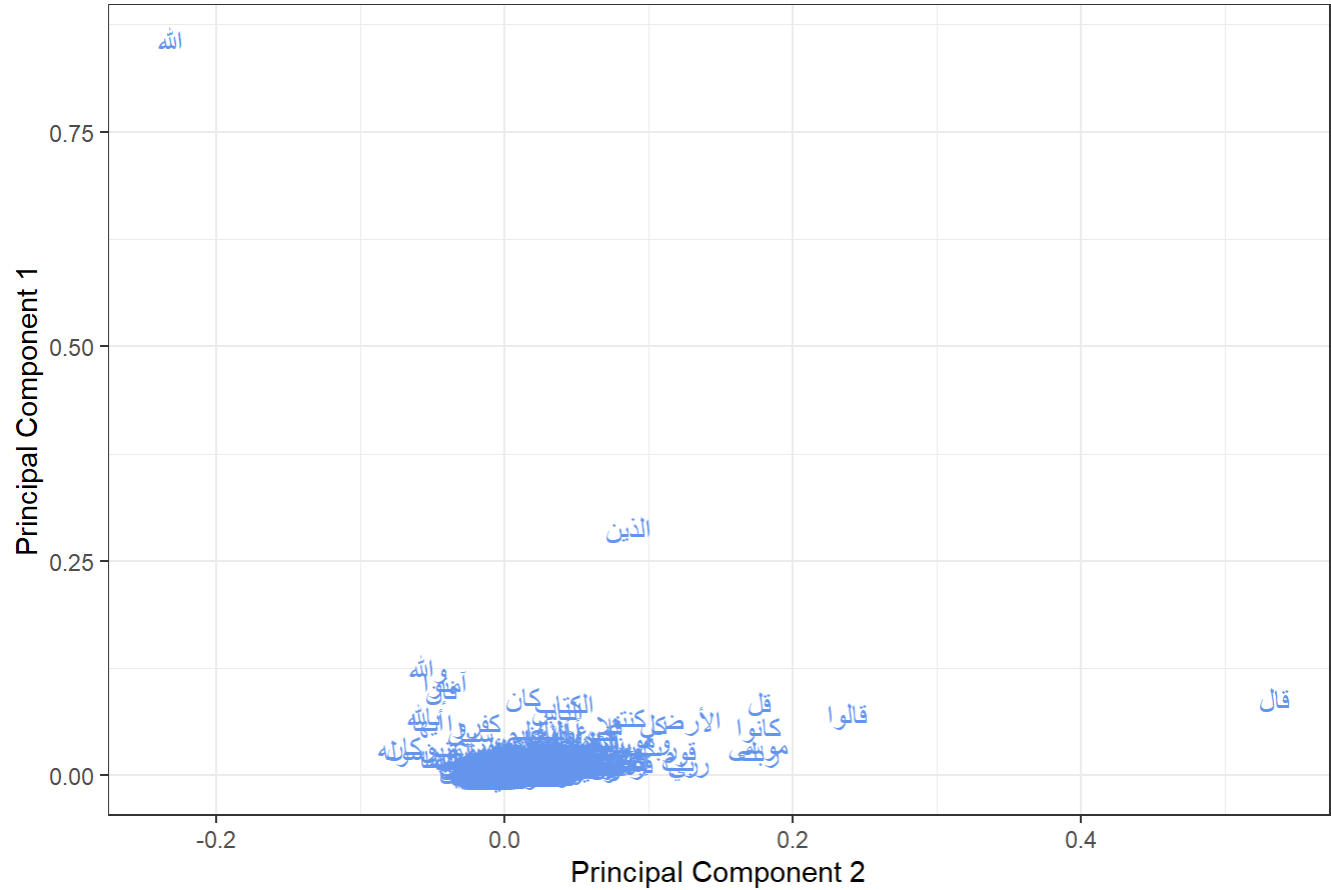


*#To Detect multidimensional outliers, if any, from the projection of
#the data:*

```
pc12=data.frame(pca1$rotation[,c(1,2)])

ggplot(pc12,
  aes(x=PC2,
    y=PC1)) +
  geom_point(color="white") +
  theme_bw() +
  ggtitle('First 2 prinipal components') +
  xlab("Principal Component 2") +
  ylab("Principal Component 1")+
  geom_text(label=rownames(pc12),
    nudge_x = 0.005,
    nudge_y = 0.005,
    size=4,colour="cornflowerblue")
```

First 2 prinipal components



```
book_id=rep(1,nrow(quran_en_sahih))
quran_en_sahih1=cbind(book_id,quran_en_sahih)
head(quran_en_sahih1)
```

##	book_id	surah_id	ayah_id	surah_title_ar	surah_title_en	surah_title_en_trans
## 1	1	1	1	الفاتحة	Al-Faatiha	The Opening
## 2	1	1	2	الفاتحة	Al-Faatiha	The Opening
## 3	1	1	3	الفاتحة	Al-Faatiha	The Opening
## 4	1	1	4	الفاتحة	Al-Faatiha	The Opening
## 5	1	1	5	الفاتحة	Al-Faatiha	The Opening
## 6	1	1	6	الفاتحة	Al-Faatiha	The Opening
##	revelation_type					
## 1	Meccan					
## 2	Meccan					
## 3	Meccan					
## 4	Meccan					
## 5	Meccan					
## 6	Meccan					
##	text surah					
## 1	In the name of Allah, the Entirely Merciful, the Especially Merciful.					1
## 2	[All] praise is [due] to Allah, Lord of the worlds -					1
## 3	The Entirely Merciful, the Especially Merciful,					1
## 4	Sovereign of the Day of Recompense.					1
## 5	It is You we worship and You we ask for help.					1
## 6	Guide us to the straight path -					1

```
## ayah ayah_title juz manzil page hizb_quarter sajda sajda_id sajda_recommended
## 1 1 1:1 1 1 1 1 FALSE NA NA
## 2 2 1:2 1 1 1 1 FALSE NA NA
## 3 3 1:3 1 1 1 1 FALSE NA NA
## 4 4 1:4 1 1 1 1 FALSE NA NA
## 5 5 1:5 1 1 1 1 FALSE NA NA
## 6 6 1:6 1 1 1 1 FALSE NA NA
## sajda_obligatory
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
```

```
quran_words <- quran_en_sahih1 %>%
  unnest_tokens(word, text) %>%
  count(book_id, word, sort = TRUE)

colnames(quran_words)=c("book_id","word","freq_in_book")

head(quran_words)
```

```
## book_id word freq_in_book
## 1 1 and 9240
## 2 1 the 7656
## 3 1 you 4570
## 4 1 of 4501
## 5 1 is 3501
## 6 1 to 3464
```

```
chapter_words <- quran_en_sahih1 %>%
  unnest_tokens(word, text) %>%
  count(surah_title_en, word, sort = TRUE)

colnames(chapter_words)=c("surah_title_en","word","freq_in_chapter")

chapter_words=na.omit(chapter_words)
head(chapter_words)
```

```
## surah_title_en word freq_in_chapter
## 1 Al-Baqara and 762
## 2 Al-Baqara the 523
## 3 An-Nisaa and 460
## 4 Al-Baqara you 425
## 5 Aal-i-Imraan and 410
## 6 Aal-i-Imraan the 389
```

```
quran_words=left_join(quran_words,chapter_words)
```

```
## Joining, by = "word"
```

```
quran_words=anti_join(quran_words,stop_words, by = "word")
head(quran_words)
```

```
##   book_id word freq_in_book surah_title_en freq_in_chapter
## 1      1 allah      2973      Al-Baqara      301
## 2      1 allah      2973      An-Nisaa      233
## 3      1 allah      2973 Aal-i-Imraan      230
## 4      1 allah      2973      At-Tawba      173
## 5      1 allah      2973      Al-Maaida      160
## 6      1 allah      2973 Al-An'aam      101
```

```
token_list=as.data.frame(table(chapter_words$surah_title_en,chapter_words$word))
token_list=token_list[,-3]
colnames(token_list)=c("surah_title_en","word")
head(token_list)
```

```
##   surah_title_en word
## 1      Al-Faatiha  a
## 2      Al-Baqara  a
## 3 Aal-i-Imraan  a
## 4      An-Nisaa  a
## 5      Al-Maaida  a
## 6 Al-An'aam  a
```

```
#book_words=mutate(book_words,ratio=book_words$freq_in_chapter/
#      book_words$freq_in_book)
head(quran_words)
```

```
##   book_id word freq_in_book surah_title_en freq_in_chapter
## 1      1 allah      2973      Al-Baqara      301
## 2      1 allah      2973      An-Nisaa      233
## 3      1 allah      2973 Aal-i-Imraan      230
## 4      1 allah      2973      At-Tawba      173
## 5      1 allah      2973      Al-Maaida      160
## 6      1 allah      2973 Al-An'aam      101
```

```
quran_words=left_join(token_list,quran_words,by=c("surah_title_en","word"))

quran_words[is.na(quran_words)]=0
head(quran_words)
```

```
##   surah_title_en word book_id freq_in_book freq_in_chapter
## 1      Al-Faatiha  a      0      0      0
## 2      Al-Baqara  a      0      0      0
## 3 Aal-i-Imraan  a      0      0      0
## 4      An-Nisaa  a      0      0      0
## 5      Al-Maaida  a      0      0      0
```


##	6	Al-An'aam	a	0	0	0
----	---	-----------	---	---	---	---

```
r=unique(quran_words$word)
c=unique(quran_ar_min$surah_title_en)

dword=matrix(quran_words$freq_in_chapter,nrow =nrow(quran_words)/114 ,
             ncol = 114,byrow = T,dimnames = list(r,c))

head(dword)
```

##		Al-Faatiha	Al-Baqara	Aal-i-Imraan	An-Nisaa	Al-Maaida	Al-An'aam	
##	a	0	0	0	0	0	0	
##	aad	0	0	0	0	0	0	
##	aaron	0	1	0	1	0	1	
##	abandon	0	0	1	0	0	0	
##	abandoned	0	0	0	0	0	0	
##	abased	0	0	0	0	0	0	
##		Al-A'raaf	Al-Anfaal	At-Tawba	Yunus	Hud	Yusuf	Ar-Ra'd
##	a	0	0	0	0	0	0	0
##	aad	2	0	1	0	4	0	1
##	aaron	3	0	0	1	0	0	0
##	abandon	1	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0
##	abased	0	0	0	0	0	0	0
##		An-Nahl	Al-Israa	Al-Kahf	Maryam	Taa-Haa	Al-Anbiyaa	Al-Hajj
##	a	0	0	0	0	0	0	0
##	aad	0	0	0	0	0	0	1
##	aaron	0	0	0	2	5	1	0
##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	1
##	abased	0	0	0	0	0	0	0
##		Al-Muminoon	An-Noor	Al-Furqaan	Ash-Shu'araa	An-Naml	Al-Qasas	
##	a	0	0	0	0	0	0	
##	aad	0	0	1	1	0	0	
##	aaron	1	0	1	2	0	1	
##	abandon	0	0	0	0	0	0	
##	abandoned	0	0	1	0	0	0	
##	abased	0	0	0	0	0	0	
##		Al-Ankaboot	Ar-Room	Luqman	As-Sajda	Al-Ahzaab	Saba	Faatir
##	a	0	0	0	0	0	0	0
##	aad	1	0	0	0	0	0	0
##	aaron	0	0	0	0	0	0	0
##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0
##	abased	0	0	0	0	0	0	0
##		As-Saaffaat	Saad	Az-Zumar	Ghafir	Fussilat	Ash-Shura	Az-Zukhruf
##	a	0	0	0	0	0	0	0
##	aad	0	1	0	1	2	0	0
##	aaron	2	0	0	0	0	0	0
##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0

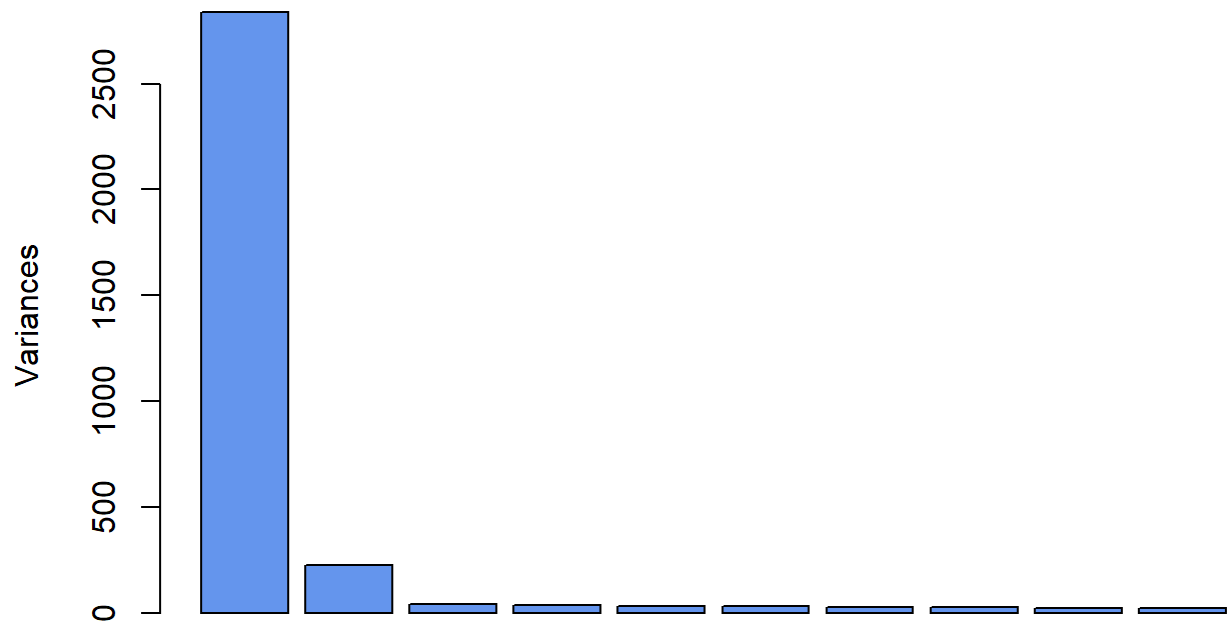
##	abased	0	0	0	0	0	0	0
##	Ad-Dukhaan	Al-Jaathiya	Al-Ahqaf	Muhammad	Al-Fath	Al-Hujuraat	Qaaf	
##	a	0	0	0	0	0	0	0
##	aad	0	0	1	0	0	0	1
##	aaron	0	0	0	0	0	0	0
##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0
##	abased	0	0	0	0	0	0	0
##	Adh-Dhaariyat	At-Tur	An-Najm	Al-Qamar	Ar-Rahmaan	Al-Waaqia	Al-Hadid	
##	a	0	0	0	0	0	0	0
##	aad	1	0	1	1	0	0	0
##	aaron	0	0	0	0	0	0	0
##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0
##	abased	0	0	0	0	0	0	0
##	Al-Mujaadila	Al-Hashr	Al-Mumtahana	As-Saff	Al-Jumu'a	Al-Munaafiqoon		
##	a	0	0	0	0	0		0
##	aad	0	0	0	0	0		0
##	aaron	0	0	0	0	0		0
##	abandon	0	0	0	0	0		0
##	abandoned	0	0	0	0	0		0
##	abased	2	0	0	0	0		0
##	At-Taghaabun	At-Talaag	At-Tahrim	Al-Mulk	Al-Qalam	Al-Haaqqa		
##	a	0	0	0	0	0		0
##	aad	0	0	0	0	0	2	
##	aaron	0	0	0	0	0	0	
##	abandon	0	0	0	0	0	0	
##	abandoned	0	0	0	0	0	0	
##	abased	0	0	0	0	0	0	
##	Al-Ma'aarij	Nooh	Al-Jinn	Al-Muzzammil	Al-Muddaththir	Al-Qiyaama		
##	a	0	0	0	0	0		0
##	aad	0	0	0	0	0		0
##	aaron	0	0	0	0	0		0
##	abandon	0	0	0	0	0		0
##	abandoned	0	0	0	0	0		0
##	abased	0	0	0	0	0		0
##	Al-Insaan	Al-Mursalaat	An-Naba	An-Naazi'aat	Abasa	At-Takwir		
##	a	0	0	0	0	0		0
##	aad	0	0	0	0	0		0
##	aaron	0	0	0	0	0		0
##	abandon	0	0	0	0	0		0
##	abandoned	0	0	0	0	0		0
##	abased	0	0	0	0	0		0
##	Al-Infitaar	Al-Mutaffifin	Al-Inshiqaaq	Al-Burooj	At-Taariq	Al-A'laa		
##	a	0	0	0	0	0		0
##	aad	0	0	0	0	0		0
##	aaron	0	0	0	0	0		0
##	abandon	0	0	0	0	0		0
##	abandoned	0	0	0	0	0		0
##	abased	0	0	0	0	0		0
##	Al-Ghaashiya	Al-Fajr	Al-Balad	Ash-Shams	Al-Lail	Ad-Dhuhaa	Ash-Sharh	
##	a	0	0	0	0	0	0	0
##	aad	0	1	0	0	0	0	0
##	aaron	0	0	0	0	0	0	0

##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0
##	abased	0	0	0	0	0	0	0
##		At-Tin	Al-Alaq	Al-Qadr	Al-Bayyina	Az-Zalzala	Al-Aadiyaat	Al-Qaari'a
##	a	0	0	0	0	0	0	0
##	aad	0	0	0	0	0	0	0
##	aaron	0	0	0	0	0	0	0
##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0
##	abased	0	0	0	0	0	0	0
##		At-Takaathur	Al-Asr	Al-Humaza	Al-Fil	Quraish	Al-Maa'un	Al-Kawthar
##	a	0	0	0	0	0	0	0
##	aad	0	0	0	0	0	0	0
##	aaron	0	0	0	0	0	0	0
##	abandon	0	0	0	0	0	0	0
##	abandoned	0	0	0	0	0	0	0
##	abased	0	0	0	0	0	0	0
##		Al-Kaafiroon	An-Nasr	Al-Masad	Al-Ikhlaas	Al-Falaq	An-Naas	
##	a	0	0	0	0	0	0	
##	aad	0	0	0	0	0	0	
##	aaron	0	0	0	0	0	0	
##	abandon	0	0	0	0	0	0	
##	abandoned	0	0	0	0	0	0	
##	abased	0	0	0	0	0	0	

```
pca1=prcomp(t(dword))

screepLOT(pca1,col="cornflowerblue",main = "Scree plot of pca word distn in englis.")
```

Scree plot of pca word distn in englis.



```
#To rank the we sort with respect to first principal components:
ranking=pca1[[2]][order(pca1[[2]][,1],decreasing = TRUE),]
head(ranking[,1],10)
```

##	allah	people	lord	fear	believed	punishment	messenger
##	0.91714760	0.16646721	0.12243153	0.08556685	0.07403904	0.06335479	0.06208374
##	knowing	believers	earth				
##	0.05850641	0.05825003	0.05814894				

```
tail(ranking[,1],10)
```

##	honored	reclining	flame	spring	blast
##	-0.0004713838	-0.0004885815	-0.0005332774	-0.0005362748	-0.0006023991
##	proportioned	burn	warning	deniers	favours
##	-0.0006236692	-0.0007638654	-0.0009697225	-0.0010271081	-0.0012893574

```
tail1=tail(ranking,20)
ranked=rownames(ranking)
```

```
#After ranking we got in top ranked words in English:
head(ranked,100)
```

##	[1]	"allah"	"people"	"lord"	"fear"	"believed"
----	-----	---------	----------	--------	--------	------------

##	[6]	"punishment"	"messenger"	"knowing"	"believers"	"earth"
##	[11]	"day"	"muhammad"	"scripture"	"revealed"	"truth"
##	[16]	"disbelievers"	"disbelieve"	"fire"	"women"	"prayer"
##	[21]	"reward"	"merciful"	"wills"	"righteous"	"religion"
##	[26]	"hearts"	"life"	"verses"	"heavens"	"remember"
##	[31]	"moses"	"wise"	"abraham"	"forgiving"	"children"
##	[36]	"spend"	"messengers"	"worship"	"guided"	"fight"
##	[41]	"mercy"	"follow"	"guidance"	"evil"	"angels"
##	[46]	"death"	"sin"	"disbelieved"	"world"	"signs"
##	[51]	"favor"	"wealth"	"blame"	"exalted"	"surely"
##	[56]	"bring"	"covenant"	"wrongdoers"	"deeds"	"disbelief"
##	[61]	"book"	"forgiveness"	"belongs"	"permission"	"abide"
##	[66]	"bounty"	"time"	"conceal"	"al"	"satan"
##	[71]	"knowledge"	"eternally"	"enter"	"hearing"	"witness"
##	[76]	"hands"	"leave"	"painful"	"wronged"	"sign"
##	[81]	"repentance"	"tidings"	"prophet"	"sufficient"	"resurrection"
##	[86]	"earned"	"lawful"	"wives"	"killed"	"paradise"
##	[91]	"land"	"kill"	"seek"	"recall"	"witnesses"
##	[96]	"israel"	"provision"	"return"	"mary"	"returned"

#After ranking we got in last ranked in words English:
tail(ranked,100)

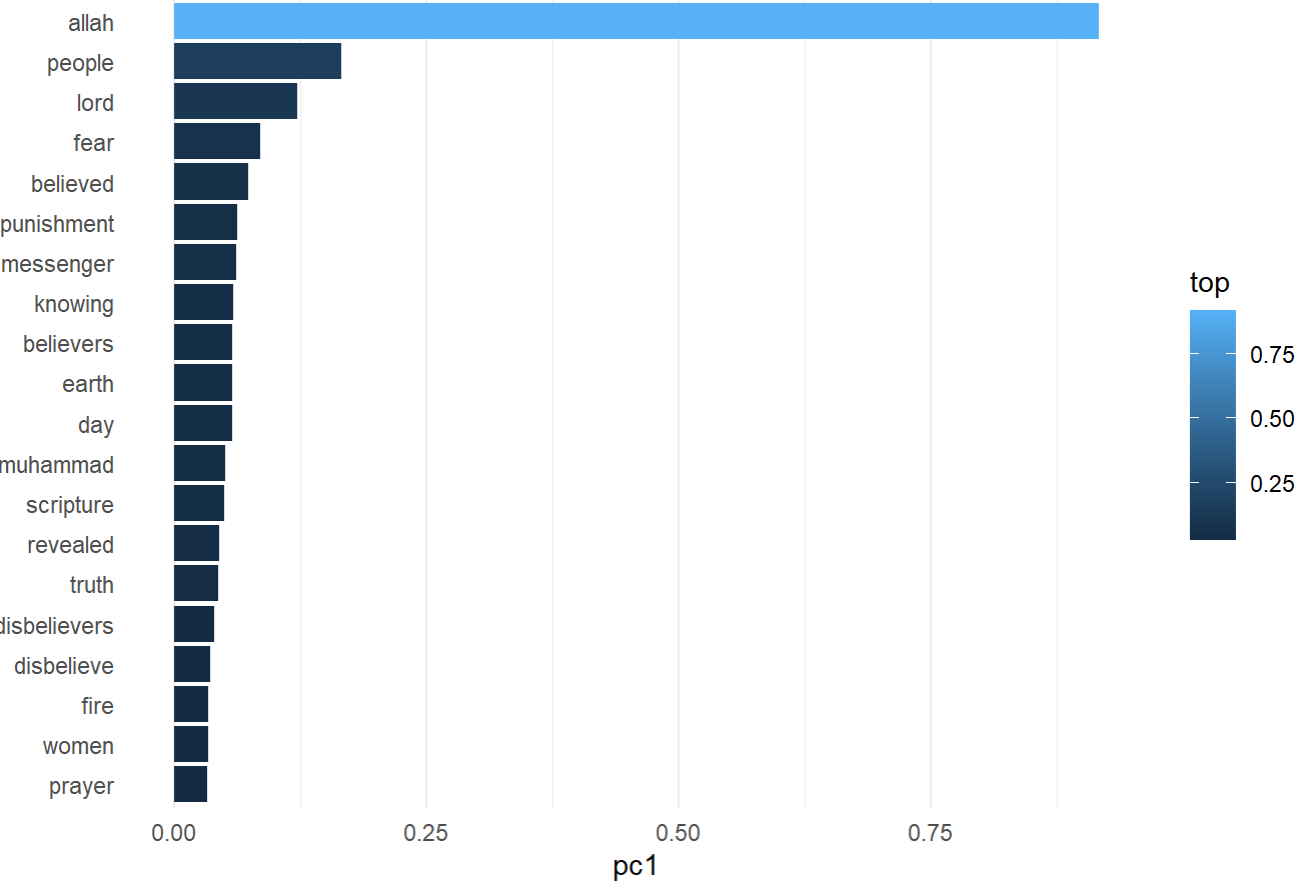
##	[1]	"prays"	"leg"	"scorner"	"zaqqum"
##	[5]	"balance"	"substance"	"multitudes"	"severely"
##	[9]	"expecting"	"comer"	"planning"	"columns"
##	[13]	"crusher"	"humanity"	"sparks"	"firewood"
##	[17]	"hot"	"disintegrated"	"swimming"	"observing"
##	[21]	"repel"	"uninterrupted"	"foul"	"strengthened"
##	[25]	"contents"	"blazing"	"pairs"	"screaming"
##	[29]	"pearls"	"walks"	"ruined"	"vessels"
##	[33]	"enumerated"	"scorching"	"inscribed"	"poet"
##	[37]	"cool"	"ahead"	"shout"	"stakes"
##	[41]	"glances"	"brocade"	"extent"	"bitter"
##	[45]	"thrones"	"deprived"	"setting"	"boys"
##	[49]	"constructed"	"mad"	"happiness"	"assembled"
##	[53]	"deliberated"	"likenesses"	"inevitable"	"sailing"
##	[57]	"barrier"	"pouring"	"obliterated"	"recorded"
##	[61]	"petitioner"	"pen"	"eternal"	"seas"
##	[65]	"lote"	"ascend"	"report"	"brightness"
##	[69]	"mixture"	"arise"	"graves"	"fair"
##	[73]	"thrown"	"preventer"	"pass"	"adorned"
##	[77]	"saqar"	"generous"	"striking"	"sovereign"
##	[81]	"occur"	"smoke"	"scriptures"	"beautiful"
##	[85]	"calamity"	"recitation"	"certainty"	"cup"
##	[89]	"couches"	"madman"	"honored"	"reclining"
##	[93]	"flame"	"spring"	"blast"	"proportioned"
##	[97]	"burn"	"warning"	"deniers"	"favors"

#To plot Top 20 words in the ranking

```
ranking1=data.frame(x1=rownames(ranking)[1:20], top=ranking[1:20,1])
```

```
ggplot(ranking1, aes(x = fct_rev(reorder(x1,-top)),
                    y = top, fill = top)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Top 20 words in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

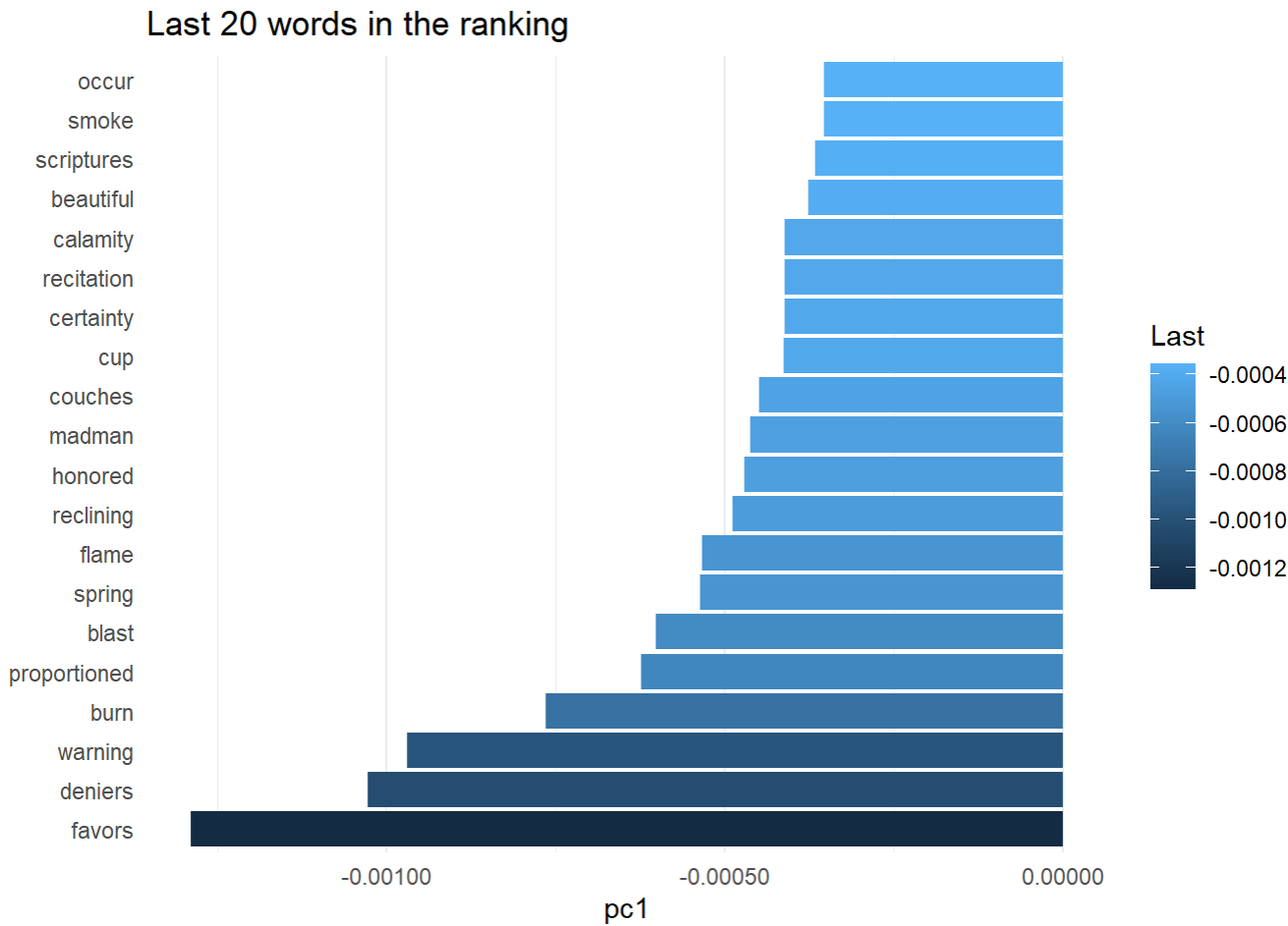
Top 20 words in the ranking



```
#To plot Last 20 words in the ranking

ranking2=data.frame(x1=rownames(tail1),Last=tail1[,1])

ggplot(ranking2, aes(x = fct_rev(reorder(x1,-Last)),
                    y = Last, fill = Last)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Last 20 words in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
dword1=as.data.frame(dword)

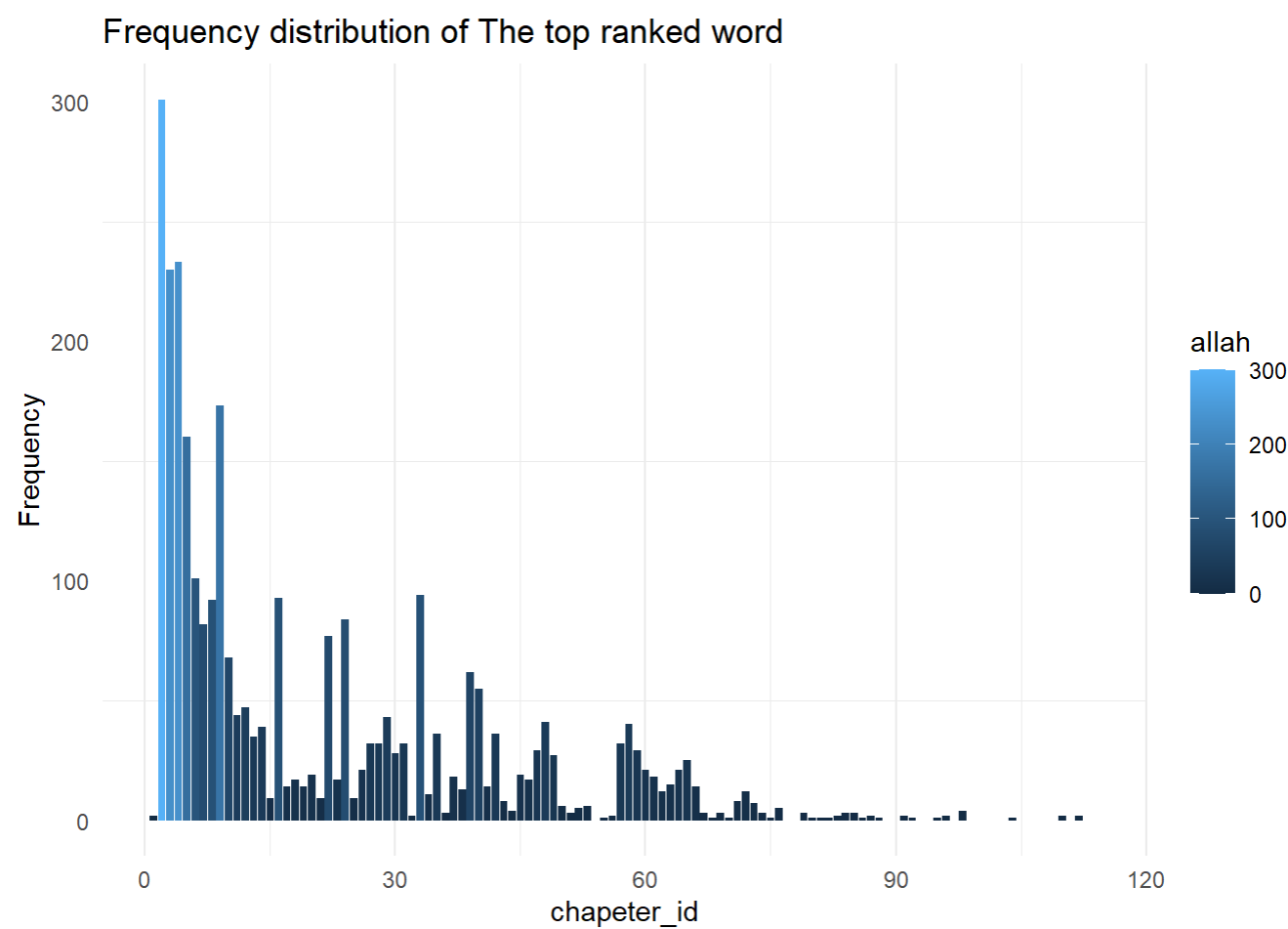
f=as.matrix(filter(dword1 ,rownames(dword1) %in% head(ranked,1)))
fl=as.matrix(filter(dword1 ,rownames(dword1) %in% tail(ranked,1)))

#view(t(f))
#view(t(fl))

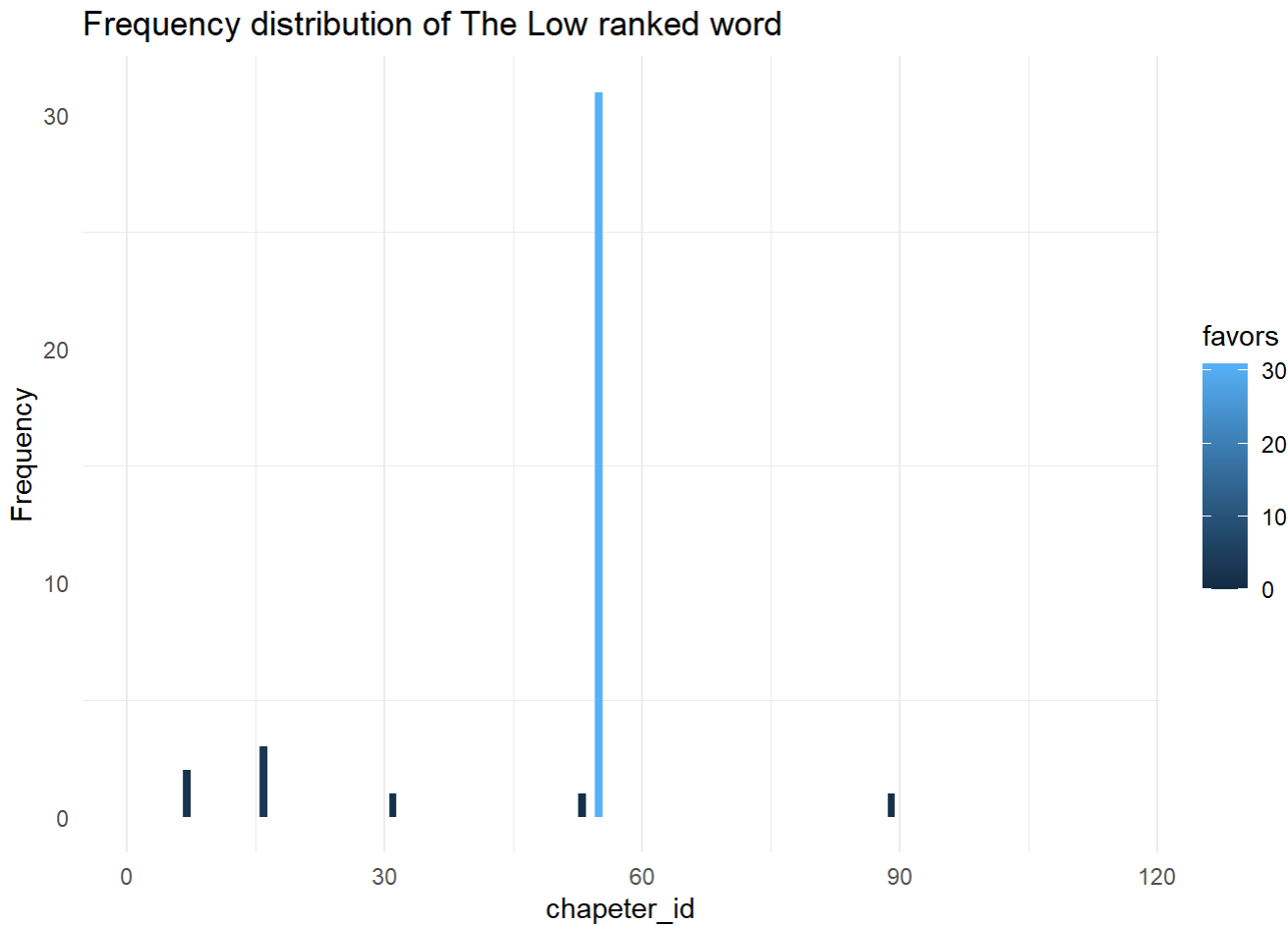
f=as.data.frame(t(f))
f=cbind(f,chapter_id=1:114)

fl=as.data.frame(t(fl))
fl=cbind(fl,chapter_id=1:114)

#Frequency distribution of The top ranked word
ggplot(f,aes(fill=allah,x=chapter_id,y=allah))+
  geom_col()+
  labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The top ranked word") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
#Frequency distribution of The Low ranked word
ggplot(f1,aes(fill=favors,x=chapter_id,y=favors))+
  geom_col()+
  labs(y = "Frequency", x ="chapter_id",title = "Frequency distribution of The Low ranked word") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

#Manually ranking i.e univariately:

```
dword11=data.frame(cbind(dword,rowSums(dword)))
dword11$zeros=rowSums(dword11==0)
dword11$occur=114-(dword11$zeros)
dword11$combine=(.4*dword11$V115)+(.6*dword11$occur)
write.table(dword11, file = 'posts.csv',sep = "\t", fileEncoding = "UTF-8")
#=dw
write.table(dword11,row.names = T,file = "wordfreq1.csv",sep = "\t",fileEncoding = "UTF-8")

ranked=rownames(dword11[order(dword11$combine, decreasing = TRUE),])

head(ranked)
```

```
## [1] "allah"      "lord"      "people"    "day"      "earth"
## [6] "punishment"
```

tail(ranked)

```
## [1] "you"        "young"     "your"      "yours"     "yourself"
## [6] "yourselves"
```

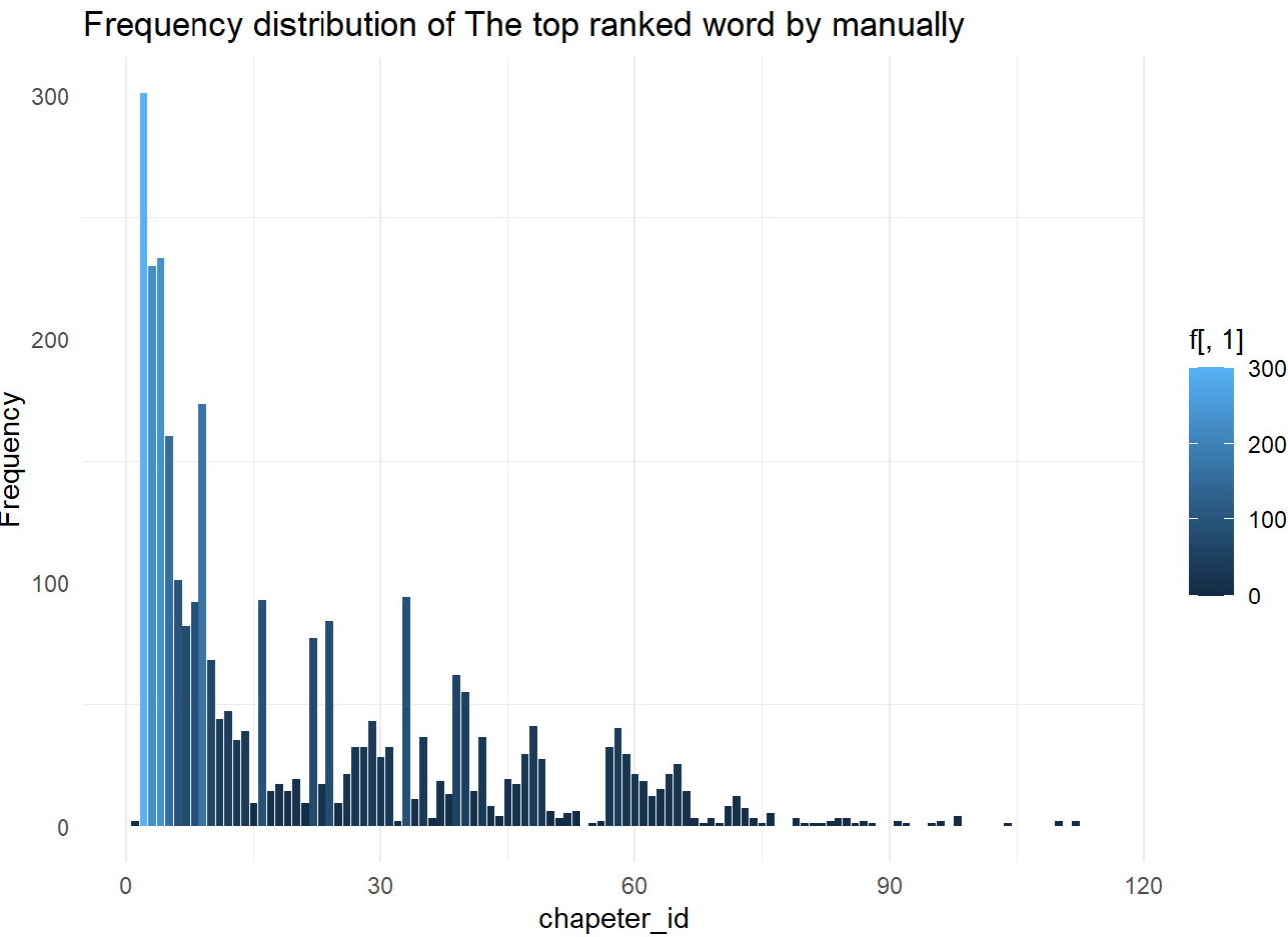
```
dword13=as.data.frame(dword11)
```

```
f=as.matrix(filter(dword1 ,rownames(dword1) %in% head(ranked,1)))
f1=as.matrix(filter(dword1 ,rownames(dword1) %in% tail(ranked,1)))

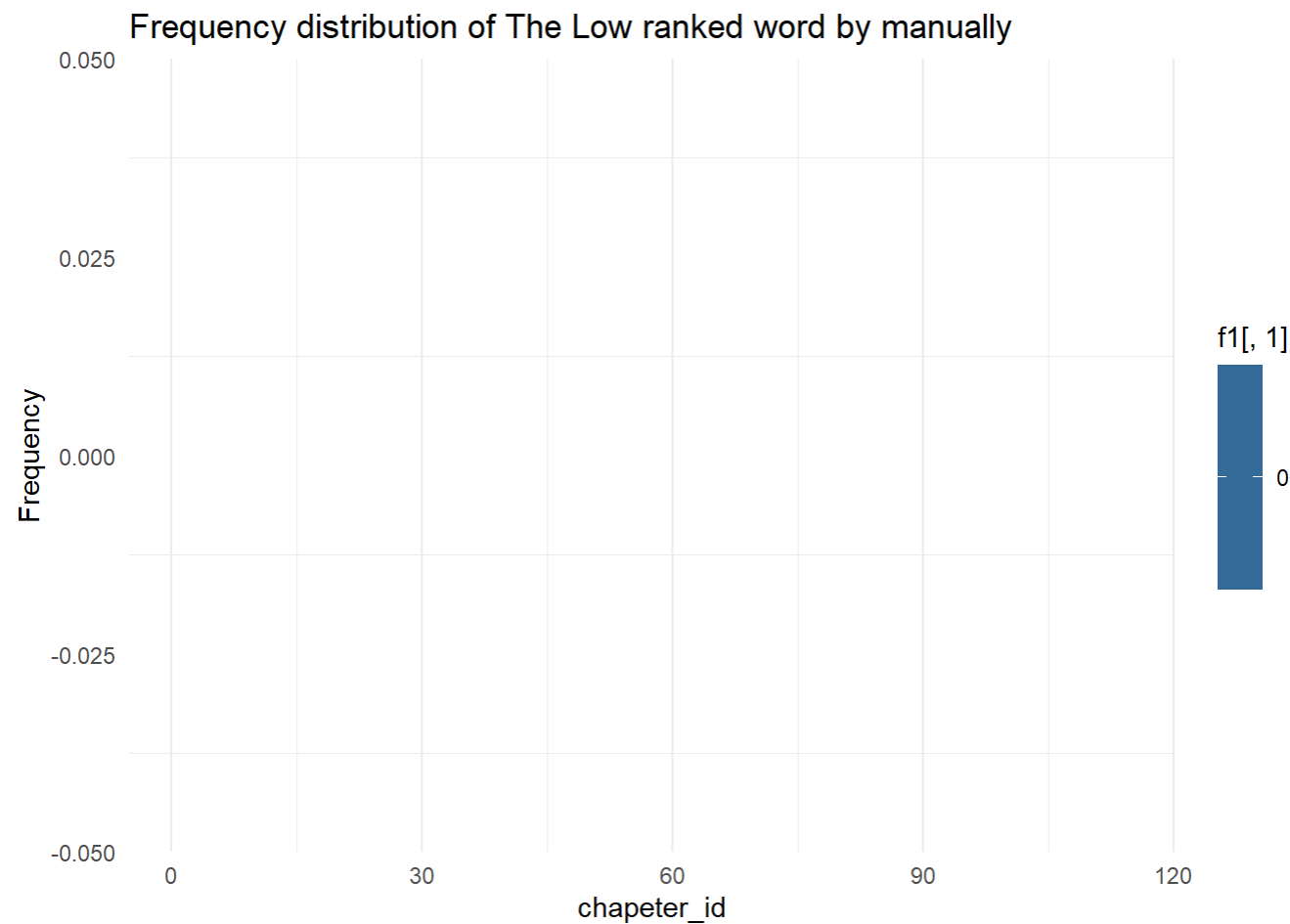
f=as.data.frame(t(f))
f=cbind(f,chapter_id=1:114)

f1=as.data.frame(t(f1))
f1=cbind(f1,chapter_id=1:114)

#Frequency distribution of The top ranked word
ggplot(f,aes(fill=f[,1],x=chapter_id,y=f[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The top ranked word by manually") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
#Frequency distribution of The Low ranked word
ggplot(f1,aes(fill=f1[,1],x=chapter_id,y=f1[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The Low ranked word by manually") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

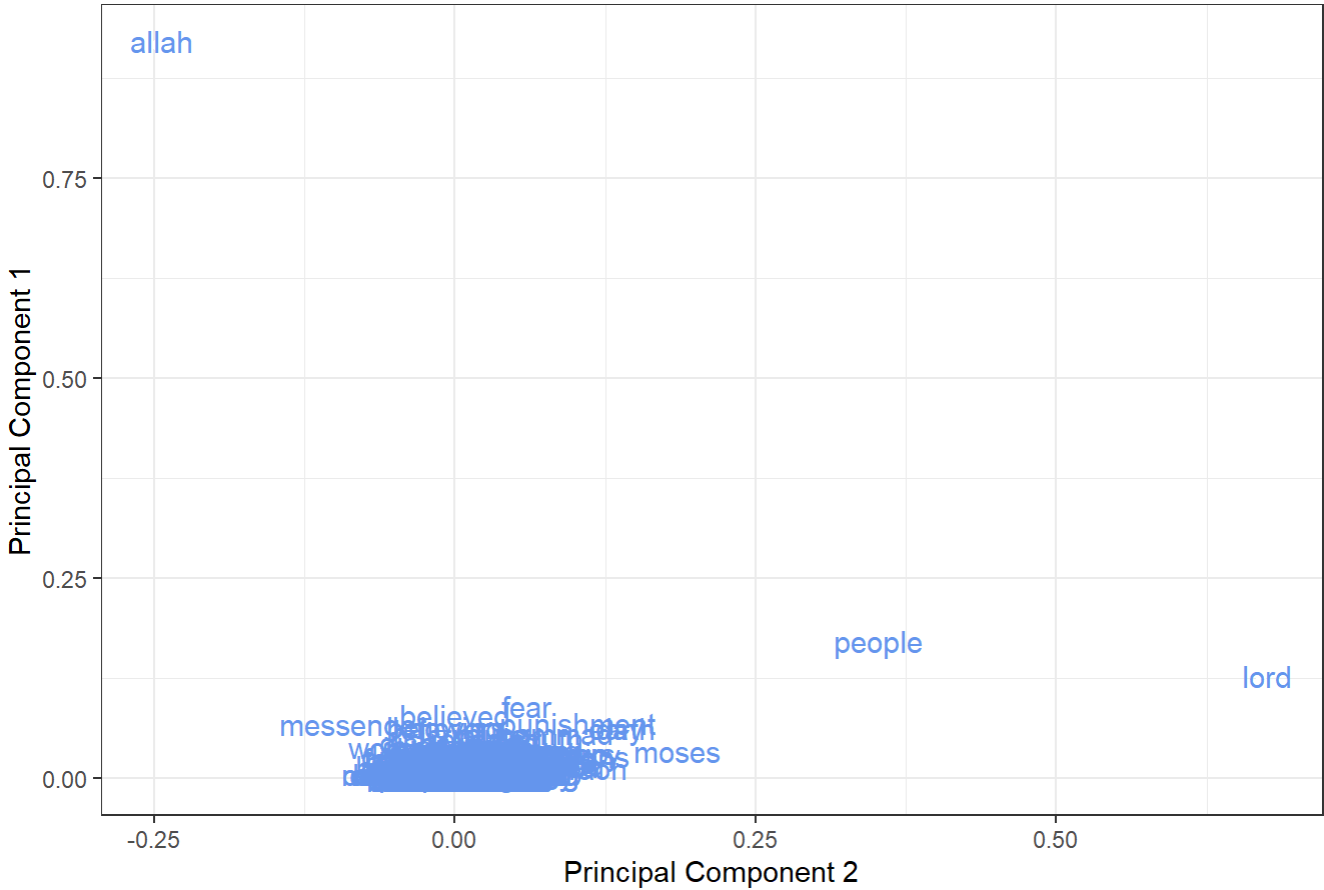


*#To Detect multidimensional outliers, if any, from the projection of
#the data:*

```
pc12=data.frame(pca1$rotation[,c(1,2)])

ggplot(pc12,
  aes(x=PC2,
    y=PC1)) +
  geom_point(color="white") +
  theme_bw() +
  ggtitle('First 2 prinipal components') +
  xlab("Principal Component 2") +
  ylab("Principal Component 1")+
  geom_text(label=rownames(pc12),
    nudge_x = 0.005,
    nudge_y = 0.005,
    size=4,colour="cornflowerblue")
```

First 2 prinipal components



#To ranke of the Holy Qur'an bigrams in arabic:

#####

```
book_id=rep(1,nrow(quran_ar_min))
quran_ar_min1=cbind(book_id,quran_ar_min)
head(quran_ar_min1)
```

##	book_id	surah_id	ayah_id	surah_title_ar	surah_title_en	surah_title_en_trans		
## 1	1	1	1	الفاتحة	Al-Faatiha	The Opening		
## 2	1	1	2	الفاتحة	Al-Faatiha	The Opening		
## 3	1	1	3	الفاتحة	Al-Faatiha	The Opening		
## 4	1	1	4	الفاتحة	Al-Faatiha	The Opening		
## 5	1	1	5	الفاتحة	Al-Faatiha	The Opening		
## 6	1	1	6	الفاتحة	Al-Faatiha	The Opening		
##	revelation_type	text			surah	ayah	ayah_title	juz
## 1	Meccan	<U+FEFF>1			1:1	1	1	بسم الله الرحمن الرحيم
## 2	Meccan		1	1:2	2	1		الحمد لله رب العالمين
## 3	Meccan			1	1:3	3	1	الرحمن الرحيم
## 4	Meccan			1	1:4	4	1	مالك يوم الدين
## 5	Meccan		1	1:5	5	1		إياك نعبد وإياك نستعين
## 6	Meccan		1	1:6	6	1		اهدنا الصراط المستقيم
##	manzil	page	hizb	quarter	sajda	sajda_id	sajda_recommended	sajda_obligatory
## 1	1	1		1	FALSE	NA	NA	NA
## 2	1	1		1	FALSE	NA	NA	NA
## 3	1	1		1	FALSE	NA	NA	NA

```
## 4 1 1 1 FALSE NA NA NA
## 5 1 1 1 FALSE NA NA NA
## 6 1 1 1 FALSE NA NA NA
```

```
quran_words <- quran_ar_min1 %>%
  unnest_tokens(bigram, text, token = "ngrams", n = 2) %>%
  count(book_id, bigram, sort = TRUE)

colnames(quran_words)=c("book_id","bigram","freq_in_book")

head(quran_words)
```

```
## book_id bigram freq_in_book
## 1 1 205 إن الله
## 2 1 184 الذين آمنوا
## 3 1 176 في الأرض
## 4 1 142 يا أيها
## 5 1 134 الذين كفروا
## 6 1 133 السماوات والأرض
```

```
chapter_words <- quran_ar_min1 %>%
  unnest_tokens(bigram, text, token = "ngrams", n = 2) %>%
  count(surah_title_en, bigram, sort = TRUE)

colnames(chapter_words)=c("surah_title_en","bigram","freq_in_chapter")

chapter_words=na.omit(chapter_words)
head(chapter_words)
```

```
## surah_title_en bigram freq_in_chapter
## 1 Ar-Rahmaan 31 آلاء ربكما
## 2 Ar-Rahmaan 31 ربكما تكذبان
## 3 Ar-Rahmaan 31 فبأي آلاء
## 4 An-Nisaa 25 إن الله
## 5 Al-Baqara 23 إن الله
## 6 Al-Baqara 21 الذين آمنوا
```

```
quran_words=left_join(quran_words,chapter_words)
```

```
## Joining, by = "bigram"
```

```
#book_words=anti_join(book_words,arabic_stopwords, by = "word")
head(quran_words)
```

```
## book_id bigram freq_in_book surah_title_en freq_in_chapter
## 1 1 205 إن الله An-Nisaa 25
## 2 1 205 إن الله Al-Baqara 23
## 3 1 205 إن الله Aal-i-Imraan 15
```

## 4	1	205	إِنَّ اللَّهَ	At-Tawba	15
## 5	1	205	إِنَّ اللَّهَ	Al-Maaida	14
## 6	1	205	إِنَّ اللَّهَ	Al-Hajj	14

```
token_list=as.data.frame(table(chapter_words$surah_title_en,chapter_words$bigram))
token_list=token_list[,-3]
colnames(token_list)=c("surah_title_en","bigram")
head(token_list)
```

```
## surah_title_en bigram
## 1 Al-Faatiha ن وَالْقَلَم
## 2 Al-Baqara ن وَالْقَلَم
## 3 Aal-i-Imraan ن وَالْقَلَم
## 4 An-Nisaa ن وَالْقَلَم
## 5 Al-Maaida ن وَالْقَلَم
## 6 Al-An'aam ن وَالْقَلَم
```

```
#book_words=mutate(book_words,ratio=book_words$freq_in_chapter/
# book_words$freq_in_book)
head(quran_words)
```

##	book_id	bigram	freq_in_book	surah_title_en	freq_in_chapter
## 1	1	205	إِنَّ اللَّهَ	An-Nisaa	25
## 2	1	205	إِنَّ اللَّهَ	Al-Baqara	23
## 3	1	205	إِنَّ اللَّهَ	Aal-i-Imraan	15
## 4	1	205	إِنَّ اللَّهَ	At-Tawba	15
## 5	1	205	إِنَّ اللَّهَ	Al-Maaida	14
## 6	1	205	إِنَّ اللَّهَ	Al-Hajj	14

```
quran_words=left_join(token_list,quran_words,by=c("surah_title_en","bigram"))

quran_words[is.na(quran_words)]=0
head(quran_words)
```

##	surah_title_en	bigram	book_id	freq_in_book	freq_in_chapter
## 1	Al-Faatiha	0	0	0	ن وَالْقَلَم
## 2	Al-Baqara	0	0	0	ن وَالْقَلَم
## 3	Aal-i-Imraan	0	0	0	ن وَالْقَلَم
## 4	An-Nisaa	0	0	0	ن وَالْقَلَم
## 5	Al-Maaida	0	0	0	ن وَالْقَلَم
## 6	Al-An'aam	0	0	0	ن وَالْقَلَم

```
r=unique(quran_words$bigram)
c=unique(quran_ar_min$surah_title_en)

dbigram=matrix(quran_words$freq_in_chapter,nrow =nrow(quran_words)/114 ,
ncol = 114,byrow = T,dimnames = list(r,c))
```

head(dbigram)

##	Al-Faatiha Al-Baqara Aal-i-Imraan An-Nisaa Al-Maaida Al-An'aam					
## 0	0	0	0	0	0	ن والقلم
## 0	1	1	0	2	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 1	0	0	0	0	0	أن لم
##	Al-A'raaf Al-Anfaal At-Tawba Yunus Hud Yusuf Ar-Ra'd Ibrahim Al-Hijr					
## 0	0	0	0	0	0	ن والقلم
## 0	0	1	1	0	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	An-Nahl Al-Israa Al-Kahf Maryam Taa-Haa Al-Anbiyaa Al-Hajj Al-Muminoon					
## 0	0	0	0	0	0	ن والقلم
## 0	2	0	0	3	1	له من
## 0	1	0	0	0	0	له إن
## 0	1	1	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	An-Noor Al-Furqaan Ash-Shu'araa An-Naml Al-Qasas Al-Ankaboot Ar-Room					
## 0	0	0	0	0	0	ن والقلم
## 1	0	1	0	0	0	له من
## 0	1	0	0	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	Luqman As-Sajda Al-Ahzaab Saba Faatir Yaseen As-Saaffaat Saad Az-Zumar					
## 0	0	0	0	0	0	ن والقلم
## 4	1	0	0	1	0	له من
## 0	0	0	1	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	Ghafir Fussilat Ash-Shura Az-Zukhruf Ad-Dukhaan Al-Jaathiya Al-Ahqaf					
## 0	0	0	0	0	0	ن والقلم
## 1	0	0	1	3	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	1	من لم
## 0	0	0	0	0	0	أن لم
##	Muhammad Al-Fath Al-Hujuraat Qaaf Adh-Dhaariyat At-Tur An-Najm					
## 0	0	0	0	0	0	ن والقلم
## 0	1	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	1	1	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	Al-Qamar Ar-Rahmaan Al-Waaqia Al-Hadid Al-Mujaadila Al-Hashr					
## 0	0	0	0	0	0	ن والقلم

## 0	0	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	Al-Mumtahana As-Saff Al-Jumu'a Al-Munaafiqoon At-Taghaabun At-Talaaq					
## 0	0	0	0	0	0	ن والقلم
## 1	0	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 0	1	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	At-Tahrim Al-Mulk Al-Qalam Al-Haaqqa Al-Ma'aarij Nooh Al-Jinn					
## 0	0	0	0	1	0	ن والقلم
## 0	0	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 3	0	0	0	0	0	أن لن
## 0	1	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	Al-Muzzammil Al-Muddaththir Al-Qiyaama Al-Insaan Al-Mursalaat An-Naba					
## 0	0	0	0	0	0	ن والقلم
## 0	0	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	0	1	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	An-Naazi'aat Abasa At-Takwir Al-Infitaar Al-Mutaffifin Al-Inshiqaaq					
## 0	0	0	0	0	0	ن والقلم
## 0	0	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 1	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	Al-Burooj At-Taariq Al-A'laa Al-Ghaashiya Al-Fajr Al-Balad Ash-Shams					
## 0	0	0	0	0	0	ن والقلم
## 0	0	0	0	0	1	له من
## 0	0	0	0	0	0	له إن
## 0	1	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	1	0	0	0	0	أن لم
##	Al-Lail Ad-Dhuhaa Ash-Sharh At-Tin Al-Alaq Al-Qadr Al-Bayyina					
## 0	0	0	0	0	0	ن والقلم
## 0	0	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم
##	Az-Zalzala Al-Aadiyaat Al-Qaari'a At-Takaathur Al-Asr Al-Humaza Al-Fil					
## 0	0	0	0	0	0	ن والقلم
## 0	0	0	0	0	0	له من
## 0	0	0	0	0	0	له إن
## 0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	أن لم

##	Quraish	Al-Maa'un	Al-Kawthar	Al-Kaafiroon	An-Nasr	Al-Masad	Al-Ikhlaas
## 0	0	0	0	0	0	0	ن والقلم
## 0	0	0	0	0	0	0	له من
## 0	0	0	0	0	0	0	له إن
## 0	0	0	0	0	0	0	أن لن
## 0	0	0	0	0	0	0	من لم
## 0	0	0	0	0	0	0	أن لم
##	Al-Falaq	An-Naas					
## 0	0		ن والقلم				
## 0	0		له من				
## 0	0		له إن				
## 0	0		أن لن				
## 0	0		من لم				
## 0	0		أن لم				

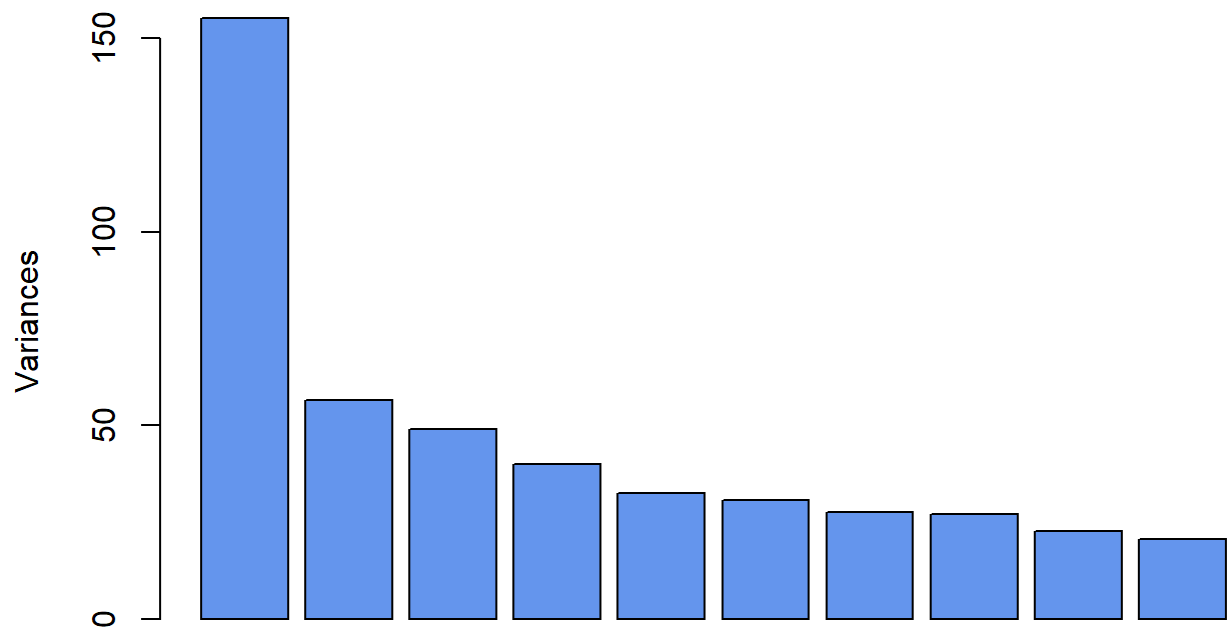
```
pca1=prcomp(t(dbigram))

pc12=data.frame(pca1$rotation[,c(1,2)])
head(pc12)
```

##	PC1	PC2
## 2.932341	ن والقلم	e-04 -1.034982e-04
## 1.752142-	له من	e-02 -4.128361e-03
## 5.977504-	له إن	e-04 -2.170245e-05
## 1.224167	أن لن	e-03 9.629559e-04
## 5.537007	من لم	e-05 -1.234800e-04
## 3.509060-	أن لم	e-04 -1.153410e-03

```
screeplot(pca1,col="cornflowerblue",main = "Scree plot of pca bigram distn
in arabic.")
```

Scree plot of pca bigram distn in arabic.



#To rank the we sort with respect to first principal components:

```
ranking=pca1[[2]][order(pca1[[2]][,1],decreasing = TRUE),]  
head(ranking[,1],10)
```

```
## فبأي آلاء آلاء ربكما ربكما تكذبان أدراك ما يومئذ للمكذبين  
## 0.011934167 0.011682093 0.011682093 0.004023305 0.003800357  
## وما أدراك ويل يومئذ يوم الدين الرحيم إذا وأما من  
## 0.003710040 0.003506749 0.002619807 0.002069079 0.001953819
```

```
tail(ranking[,1],10)
```

```
## في سبيل من الله الذين كفروا سبيل الله من بعد في الأرض  
## -0.1090083 -0.1160677 -0.1194212 -0.1285186 -0.1478060 -0.1486087  
## أيها الذين يا أيها الذين آمنوا إن الله  
## -0.1562173 -0.1901616 -0.2501936 -0.3147273
```

```
tail1=tail(ranking,20)  
ranked=rownames(ranking)
```

#After ranking we got in top ranked bigrams in Arabic:

```
head(ranked,100)
```

```
## [1] "فبأي آلاء" "آلاء ربكما" "ربكما تكذبان" "أدراك ما"
```

"يومئذ للمكذبين" "وما أدراك" "ويل يومئذ" "يوم الدين" [5] ##
 ## "الرحيم إذا" "وأما من" "من مذكر" "عذابي ونذر" [9] ##
 ## "فهل من" "رب السماوات" "إنهم كانوا" "عباد الله" [13] ##
 ## "إلى ربك" "إنا أرسلنا" "إن المتقين" "المتقين في" [17] ##
 ## "الذين هم" "العزيز الرحيم" "والليل إذا" "الإنسان ما" [21] ##
 ## "إنا كذلك" "الرحيم قل" "كلا بل" "خلقنا الإنسان" [25] ##
 ## "فلا أقسم" "يوم الفصل" "الله الرحمن" "فمن شاء" [29] ##
 ## "فأما من" "إلها آخر" "الله إلها" "الذي كنتم" [33] ##
 ## "إني لكم" "أوتي كتابه" "بل هم" "أتاك حديث" [37] ##
 ## "قبلهم من" "لنذكر فهل" "ولقد يسرنا" "القرآن للذكر" [41] ##
 ## "يسرنا القرآن" "كان عذابي" "إن علينا" "وجوه يومئذ" [45] ##
 ## "بل لا" "الله المخلصين" "كذلك نجزي" "كنتم به" [49] ##
 ## "أرسلنا عليهم" "جنات وعيون" "باسم ربك" "متنا وكنا" [53] ##
 ## "وكنا ترابا" "أن لن" "إنا نحن" "بسم الله" [57] ##
 ## "من أوتي" "اسم ربك" "أم نحن" "ما أصحاب" [61] ##
 ## "صيحة واحدة" "هل أتاك" "أم لكم" "كان في" [65] ##
 ## "أصحاب اليمين" "يا ويلنا" "على الأرائك" "ولا جان" [69] ##
 ## "في الآخرين" "فيها ولا" "الإنسان من" "كذب وتولي" [73] ##
 ## "إلى ربه" "به تكذبون" "ولقد خلقنا" "أم هم" [77] ##
 ## "كذبت ثمود" "على سرر" "في يوم" "يومهم الذي" [81] ##
 ## "أم عندهم" "من صلصال" "من فضة" "متكئين على" [85] ##
 ## "ترابا وعظاما" "بحمد ربك" "ومن الليل" "إلا عباد" [89] ##
 ## "سلام على" "عبادنا المؤمنين" "إذ نادى" "أهلكنا قبلهم" [93] ##
 ## "من عبادنا" "وأنه هو" "لكم رسول" "مع الله" [97] ##

#After ranking we got in top ranked bigrams in Arabic:

tail(ranked,100)

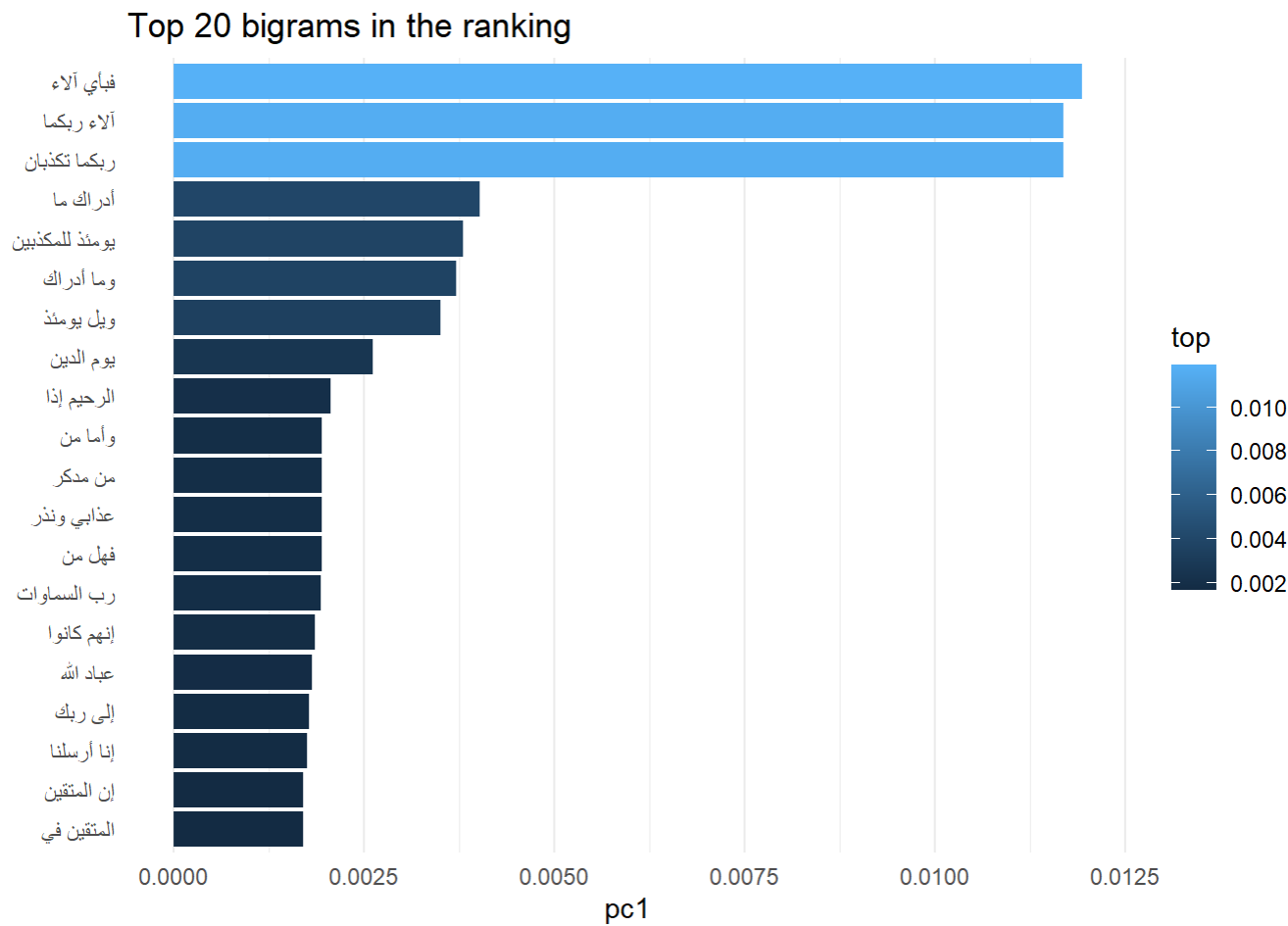
"الذين من" "ألم تر" "وما في" "السماوات وما" [1] ##
 ## "من ربهم" "الله واعلموا" "خير لكم" "الله عليكم" [5] ##
 ## "الله وما" "شاء الله" "سميع عليم" "وعملوا الصالحات" [9] ##
 ## "خوف عليهم" "هم يحزنون" "عليهم ولا" "لا يعلمون" [13] ##
 ## "ولهم عذاب" "من قبلكم" "ما لا" "يشاء والله" [17] ##
 ## "كنتم مؤمنين" "في الآخرة" "حدود الله" "الله ولا" [21] ##
 ## "هم فيها" "من تحتها" "خالدين فيها" "بعد ذلك" [25] ##
 ## "لا يحب" "والله لا" "وما كان" "تحتها الأنهار" [29] ##
 ## "أوتوا الكتاب" "تجري من" "ابن مريم" "فلا جناح" [33] ##
 ## "الله لكم" "فيها خالدون" "على الذين" "الله ورسوله" [37] ##
 ## "واعلموا أن" "عذاب أليم" "الله والله" "الله ما" [41] ##
 ## "آمنوا لا" "من الذين" "بما أنزل" "الذين أوتوا" [45] ##
 ## "بني إسرائيل" "الحياة الدنيا" "من خير" "دون الله" [49] ##
 ## "والذين آمنوا" "من ربكم" "أنزل الله" "الله غفور" [53] ##
 ## "بالله واليوم" "في السماوات" "في الدنيا" "يوم القيامة" [57] ##
 ## "شيء قدير" "واليوم الآخر" "من عند" "وما أنزل" [61] ##
 ## "بما تعملون" "إلا أن" "الله إن" "الله كان" [65] ##
 ## "ولا هم" "وكان الله" "من دون" "على الله" [69] ##
 ## "بعد ما" "غفور رحيم" "أهل الكتاب" "واتقوا الله" [73] ##
 ## "السماوات والأرض" "على كل" "كل شيء" "الله لا" [77] ##
 ## "الله على" "من قبل" "من يشاء" "الله من" [81] ##
 ## "ما في" "فإن الله" "إن الذين" "عند الله" [85] ##
 ## "إن كنتم" "أن الله" "في سبيل" "من الله" [89] ##
 ## "الذين كفروا" "سبيل الله" "من بعد" "في الأرض" [93] ##

```
## [97] "إن الله" "الذين آمنوا" "يا أيها" "أيها الذين"
```

```
#To plot Top 20 words in the ranking

ranking1=data.frame(x1=rownames(ranking)[1:20], top=ranking[1:20,1])

ggplot(ranking1, aes(x = fct_rev(reorder(x1,-top)),
  y = top, fill = top)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Top 20 bigrams in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

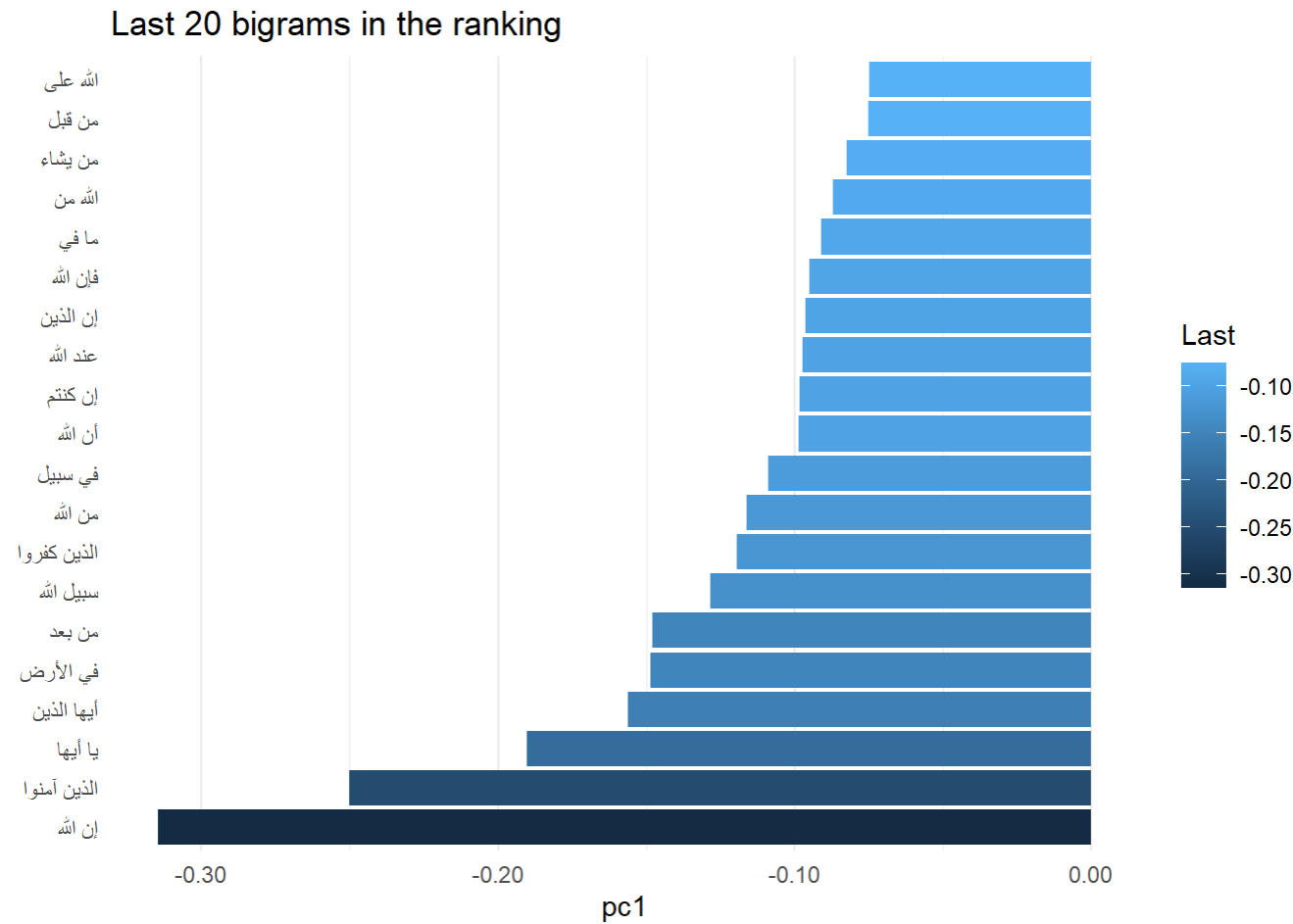


```
#To plot Last 20 words in the ranking

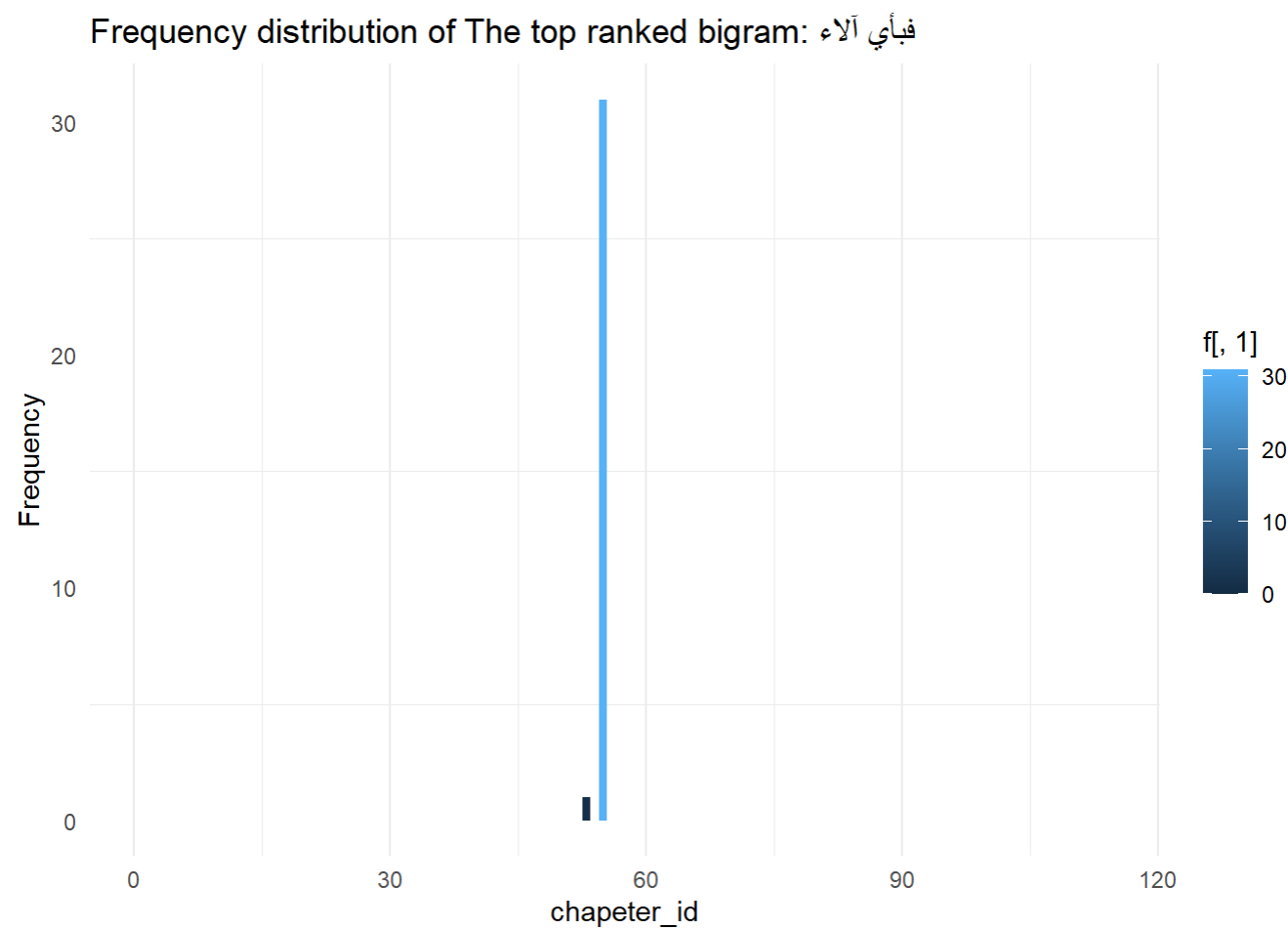
ranking2=data.frame(x1=rownames(tail1),Last=tail1[,1])

ggplot(ranking2, aes(x = fct_rev(reorder(x1,-Last)),
  y = Last, fill = Last)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Last 20 bigrams in the ranking") +
  coord_flip() +
```

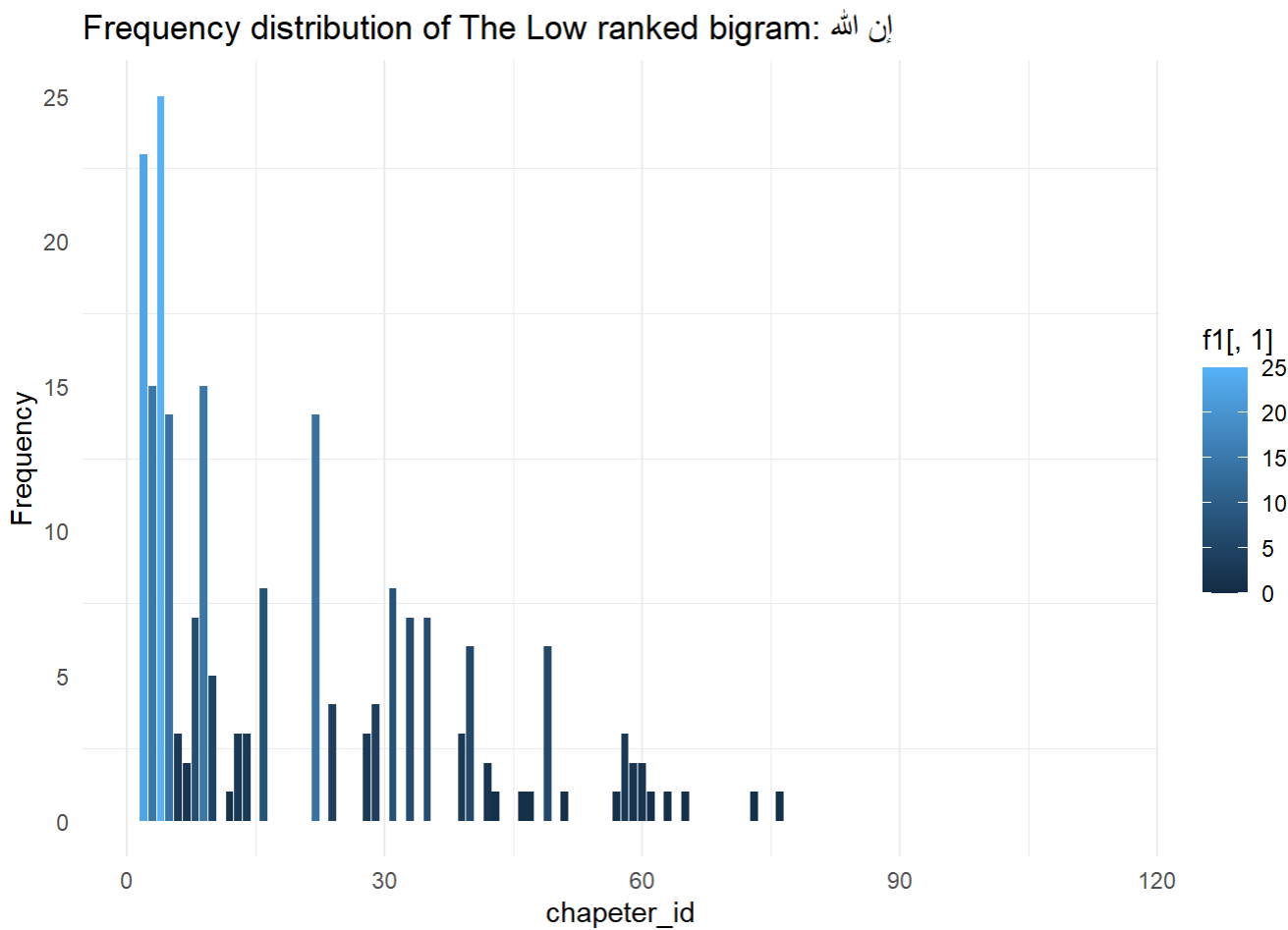
```
theme_minimal() +  
theme(panel.grid.major.y = element_blank())
```



```
dbigram1=as.data.frame(dbigram)  
  
f=as.matrix(filter(dbigram1 ,rownames(dbigram1) %in% head(ranked,1)))  
f1=as.matrix(filter(dbigram1 ,rownames(dbigram1) %in% tail(ranked,1)))  
  
#view(t(f))  
#view(t(f1))  
  
f=as.data.frame(t(f))  
f=cbind(f,chapter_id=1:114)  
  
f1=as.data.frame(t(f1))  
f1=cbind(f1,chapter_id=1:114)  
  
#Frequency distribution of The top ranked word  
ggplot(f,aes(fill=f[,1],x=chapter_id,y=f[,1]))+  
  geom_col()+  
  labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The top ranked bigram: فبأي آلاء") +  
  theme_minimal() +  
  theme(panel.grid.major.y = element_blank())
```



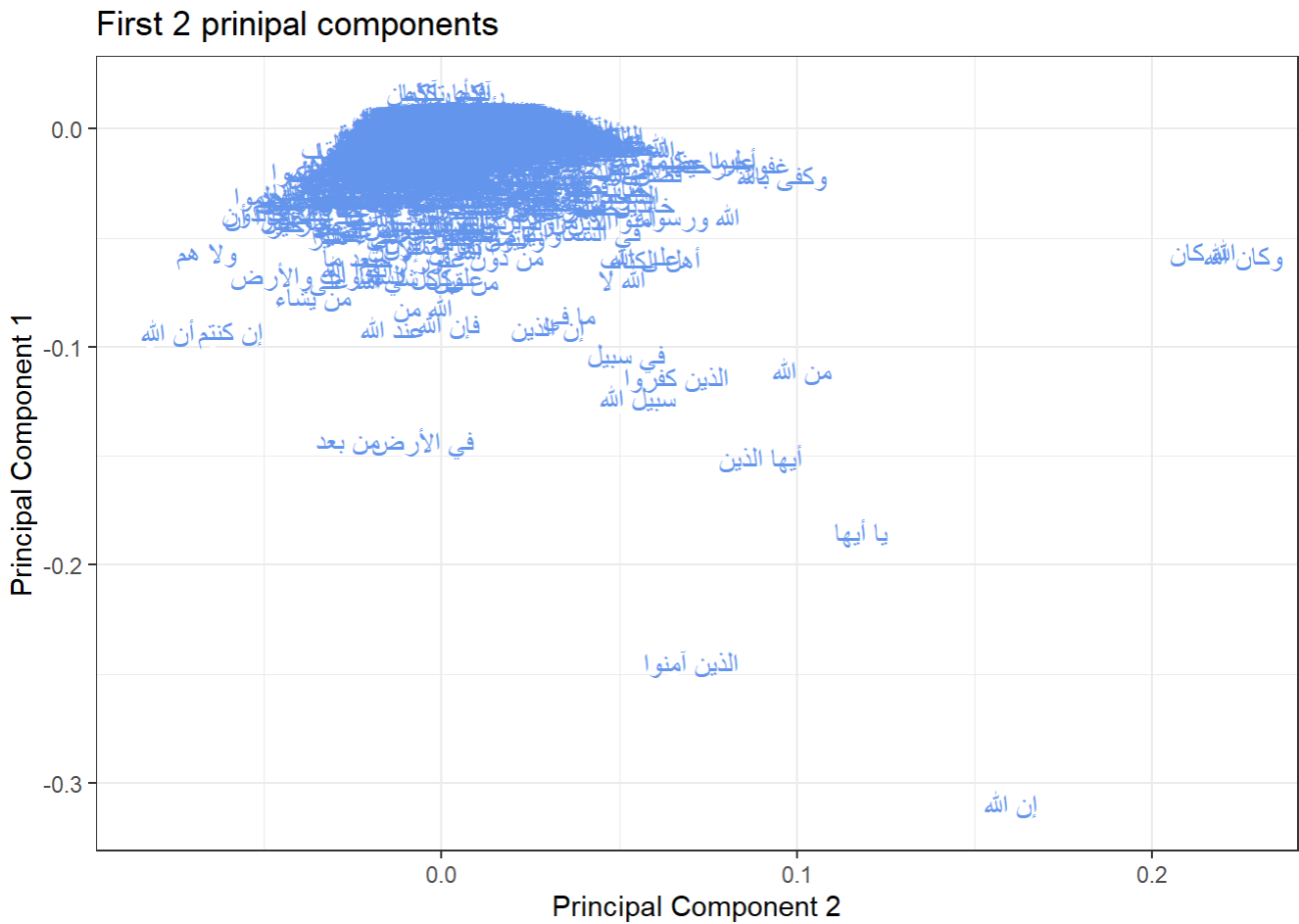
```
ggplot(f1,aes(fill=f1[,1],x=chapter_id,y=f1[,1]))+
geom_col()+
labs(y = "Frequency", x ="chapter_id",title = "Frequency distribution of The Low ranked bigram: إن الله") +
theme_minimal() +
theme(panel.grid.major.y = element_blank())
```



*#To Detect multidimensional outliers, if any, from the projection of
#the data:*

```
pc12=data.frame(pca1$rotation[,c(1,2)])

ggplot(pc12,
  aes(x=PC2,
    y=PC1)) +
  geom_point(color="white") +
  theme_bw() +
  ggtitle('First 2 prinipal components') +
  xlab("Principal Component 2") +
  ylab("Principal Component 1")+
  geom_text(label=rownames(pc12),
    nudge_x = 0.005,
    nudge_y = 0.005,
    size=4,colour="cornflowerblue")
```



#To ranke of the Holy Qur'an bigrams in english:
#####

```
book_id=rep(1,nrow(quran_en_sahih))
quran_en_sahih1=cbind(book_id,quran_en_sahih)
head(quran_en_sahih1)
```

##	book_id	surah_id	ayah_id	surah_title_ar	surah_title_en	surah_title_en_trans
## 1	1	1	1	الفاتحة	Al-Faatiha	The Opening
## 2	1	1	2	الفاتحة	Al-Faatiha	The Opening
## 3	1	1	3	الفاتحة	Al-Faatiha	The Opening
## 4	1	1	4	الفاتحة	Al-Faatiha	The Opening
## 5	1	1	5	الفاتحة	Al-Faatiha	The Opening
## 6	1	1	6	الفاتحة	Al-Faatiha	The Opening
##	revelation_type					
## 1	Meccan					
## 2	Meccan					
## 3	Meccan					
## 4	Meccan					
## 5	Meccan					
## 6	Meccan					
##	text surah					
## 1	In the name of Allah, the Entirely Merciful, the Especially Merciful.					1
## 2	[All] praise is [due] to Allah, Lord of the worlds -					1
## 3	The Entirely Merciful, the Especially Merciful,					1


```
## 4 Sovereign of the Day of Recompense. 1
## 5 It is You we worship and You we ask for help. 1
## 6 Guide us to the straight path - 1
## ayah ayah_title juz manzil page hizb_quarter sajda sajda_id sajda_recommended
## 1 1 1:1 1 1 1 1 FALSE NA NA
## 2 2 1:2 1 1 1 1 FALSE NA NA
## 3 3 1:3 1 1 1 1 FALSE NA NA
## 4 4 1:4 1 1 1 1 FALSE NA NA
## 5 5 1:5 1 1 1 1 FALSE NA NA
## 6 6 1:6 1 1 1 1 FALSE NA NA
## sajda_obligatory
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
```

```
quran_words <- quran_en_sahih1 %>%
  unnest_tokens(bigram, text, token = "ngrams", n = 2) %>%
  count(book_id, bigram, sort = TRUE)

colnames(quran_words)=c("book_id","bigram","freq_in_book")

head(quran_words)
```

```
## book_id bigram freq_in_book
## 1 1 of the 969
## 2 1 those who 900
## 3 1 and the 640
## 4 1 do not 581
## 5 1 will be 506
## 6 1 is the 496
```

```
chapter_words <- quran_en_sahih1 %>%
  unnest_tokens(bigram, text, token = "ngrams", n = 2) %>%
  count(surah_title_en, bigram, sort = TRUE)

colnames(chapter_words)=c("surah_title_en","bigram","freq_in_chapter")

chapter_words=na.omit(chapter_words)
head(chapter_words)
```

```
## surah_title_en bigram freq_in_chapter
## 1 Al-Baqara those who 79
## 2 Al-Baqara allah is 66
## 3 Aal-i-Imraan of the 56
## 4 Al-Baqara do not 55
## 5 Al-Baqara it is 55
## 6 Aal-i-Imraan those who 55
```

```
quran_words=left_join(quran_words,chapter_words)
```

```
## Joining, by = "bigram"
```

```
#arabic_stopwords = data.frame(word = removeStopWords("ÔâÇã"))$arabicStopwordList)

#book_words=anti_join(book_words,arabic_stopwords, by = "word")
head(quran_words)
```

```
##      book_id bigram freq_in_book surah_title_en freq_in_chapter
## 1          1 of the          969    Aal-i-Imraan             56
## 2          1 of the          969      Al-Baqara             50
## 3          1 of the          969    Al-A'raaf             48
## 4          1 of the          969    Al-An'aam             42
## 5          1 of the          969    Al-Maaida             39
## 6          1 of the          969    Ash-Shu'araa           37
```

```
token_list=as.data.frame(table(chapter_words$surah_title_en,chapter_words$bigram))
token_list=token_list[,-3]
colnames(token_list)=c("surah_title_en","bigram")
head(token_list)
```

```
##      surah_title_en bigram
## 1      Al-Faatiha  a bad
## 2      Al-Baqara  a bad
## 3    Aal-i-Imraan  a bad
## 4      An-Nisaa  a bad
## 5    Al-Maaida  a bad
## 6    Al-An'aam  a bad
```

```
#book_words=mutate(book_words,ratio=book_words$freq_in_chapter/
#      book_words$freq_in_book)
head(quran_words)
```

```
##      book_id bigram freq_in_book surah_title_en freq_in_chapter
## 1          1 of the          969    Aal-i-Imraan             56
## 2          1 of the          969      Al-Baqara             50
## 3          1 of the          969    Al-A'raaf             48
## 4          1 of the          969    Al-An'aam             42
## 5          1 of the          969    Al-Maaida             39
## 6          1 of the          969    Ash-Shu'araa           37
```

```
quran_words=left_join(token_list,quran_words,by=c("surah_title_en","bigram"))

quran_words[is.na(quran_words)]=0
head(quran_words)
```

```
##      surah title en bigram book id freq in book freq in chapter
```

##	1	Al-Faatiha	a bad	0	0	0
##	2	Al-Baqara	a bad	0	0	0
##	3	Aal-i-Imraan	a bad	0	0	0
##	4	An-Nisaa	a bad	0	0	0
##	5	Al-Maaida	a bad	0	0	0
##	6	Al-An'aam	a bad	0	0	0

```
r=unique(quran_words$bigram)
c=unique(quran_en_sahih$surah_title_en)

dbigram=matrix(quran_words$freq_in_chapter,nrow =nrow(quran_words)/114 ,
               ncol = 114,byrow = T,dimnames = list(r,c))

head(dbigram)
```

##		Al-Faatiha	Al-Baqara	Aal-i-Imraan	An-Nisaa	Al-Maaida	Al-An'aam	
##	a bad	0	0	0	0	0	0	
##	a band	0	0	0	0	0	0	
##	a bank	0	0	0	0	0	0	
##	a banquet	0	0	0	0	0	0	
##	a barren	0	0	0	0	0	0	
##	a barrier	0	0	0	0	0	0	
##		Al-A'raaf	Al-Anfaal	At-Tawba	Yunus	Hud	Yusuf	Ar-Ra'd Ibrahim Al-Hijr
##	a bad	1	0	0	0	0	0	2 0
##	a band	0	0	0	0	0	0	0 0
##	a bank	0	0	1	0	0	0	0 0
##	a banquet	0	0	0	0	0	1	0 0
##	a barren	0	0	0	0	0	0	0 0
##	a barrier	0	0	0	0	0	0	0 0
##		An-Nahl	Al-Israa	Al-Kahf	Maryam	Taa-Haa	Al-Anbiyaa	Al-Hajj
##	a bad	0	0	0	0	0	0	0
##	a band	0	0	0	0	0	0	0
##	a bank	0	0	0	0	0	0	0
##	a banquet	0	0	0	0	0	0	0
##	a barren	0	0	1	0	0	0	1
##	a barrier	0	0	1	0	0	0	0
##		Al-Muminoon	An-Noor	Al-Furqaan	Ash-Shu'araa	An-Naml	Al-Qasas	
##	a bad	0	0	0	0	1	0	
##	a band	0	0	0	0	0	1	
##	a bank	0	0	0	0	0	0	
##	a banquet	0	0	0	0	0	0	
##	a barren	0	0	0	0	0	0	
##	a barrier	1	0	1	0	1	0	
##		Al-Ankaboot	Ar-Room	Lugman	As-Sajda	Al-Ahzaab	Saba	Faatir Yaseen
##	a bad	0	1	0	0	0	0	1
##	a band	0	0	0	0	0	0	0
##	a bank	0	0	0	0	0	0	0
##	a banquet	0	0	0	0	0	0	0
##	a barren	0	0	0	0	0	0	0
##	a barrier	0	0	0	0	0	0	2
##		As-Saaffaat	Saad	Az-Zumar	Ghafir	Fussilat	Ash-Shura	Az-Zukhruf

##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	0	0	0	0	0	0	0
##	a	barrier	0	0	0	0	0	0	0
##		Ad-Dukhaan	Al-Jaathiya	Al-Ahqaf	Muhammad	Al-Fath	Al-Hujuraat	Qaaf	
##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	0	0	0	0	0	0	0
##	a	barrier	0	0	0	0	0	0	0
##		Adh-Dhaariyat	At-Tur	An-Najm	Al-Qamar	Ar-Rahmaan	Al-Waaqia	Al-Hadid	
##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	1	0	0	0	0	0	0
##	a	barrier	0	0	0	0	1	0	0
##		Al-Mujaadila	Al-Hashr	Al-Mumtahana	As-Saff	Al-Jumu'a	Al-Munaafiqoon		
##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	0	0	0	0	0	0	0
##	a	barrier	0	0	0	0	0	0	0
##		At-Taghaabun	At-Talaag	At-Tahrim	Al-Mulk	Al-Qalam	Al-Haaqqa		
##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	0	0	0	0	0	0	0
##	a	barrier	0	0	0	0	0	0	0
##		Al-Ma'aarij	Nooh	Al-Jinn	Al-Muzzammil	Al-Muddaththir	Al-Qiyaama		
##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	0	0	0	0	0	0	0
##	a	barrier	0	0	0	0	0	0	0
##		Al-Insaan	Al-Mursalaat	An-Naba	An-Naazi'aat	Abasa	At-Takwir		
##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	0	0	0	0	0	0	0
##	a	barrier	0	0	0	0	0	0	0
##		Al-Infitaar	Al-Mutaffifin	Al-Inshiqaaq	Al-Burooj	At-Taariq	Al-A'laa		
##	a	bad	0	0	0	0	0	0	0
##	a	band	0	0	0	0	0	0	0
##	a	bank	0	0	0	0	0	0	0
##	a	banquet	0	0	0	0	0	0	0
##	a	barren	0	0	0	0	0	0	0

##	a barrier	0	0	0	0	0	0	
##		Al-Ghaashiya	Al-Fajr	Al-Balad	Ash-Shams	Al-Lail	Ad-Dhuhaa	Ash-Sharh
##	a bad	0	0	0	0	0	0	0
##	a band	0	0	0	0	0	0	0
##	a bank	0	0	0	0	0	0	0
##	a banquet	0	0	0	0	0	0	0
##	a barren	0	0	0	0	0	0	0
##	a barrier	0	0	0	0	0	0	0
##		At-Tin	Al-Alaq	Al-Qadr	Al-Bayyina	Az-Zalzala	Al-Aadiyaat	Al-Qaari'a
##	a bad	0	0	0	0	0	0	0
##	a band	0	0	0	0	0	0	0
##	a bank	0	0	0	0	0	0	0
##	a banquet	0	0	0	0	0	0	0
##	a barren	0	0	0	0	0	0	0
##	a barrier	0	0	0	0	0	0	0
##		At-Takaathur	Al-Asr	Al-Humaza	Al-Fil	Quraish	Al-Maa'un	Al-Kawthar
##	a bad	0	0	0	0	0	0	0
##	a band	0	0	0	0	0	0	0
##	a bank	0	0	0	0	0	0	0
##	a banquet	0	0	0	0	0	0	0
##	a barren	0	0	0	0	0	0	0
##	a barrier	0	0	0	0	0	0	0
##		Al-Kaafiroon	An-Nasr	Al-Masad	Al-Ikhlaas	Al-Falaq	An-Naas	
##	a bad	0	0	0	0	0	0	
##	a band	0	0	0	0	0	0	
##	a bank	0	0	0	0	0	0	
##	a banquet	0	0	0	0	0	0	
##	a barren	0	0	0	0	0	0	
##	a barrier	0	0	0	0	0	0	

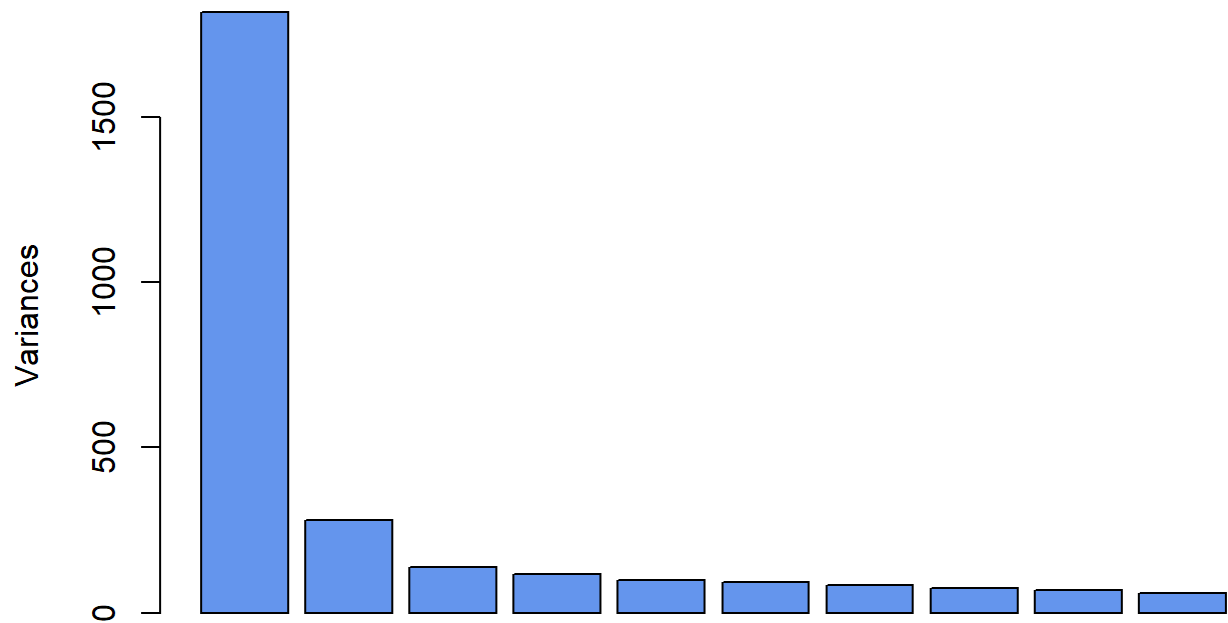
```
pca1=prcomp(t(dbigram))

pc12=data.frame(pca1$rotation[,c(1,2)])
head(pc12)
```

##		PC1	PC2
##	a bad	0.0006393044	0.0035013635
##	a band	0.0001335668	0.0008580677
##	a bank	0.0004743705	-0.0014202055
##	a banquet	0.0001542462	0.0009968137
##	a barren	0.0001857673	0.0006810080
##	a barrier	0.0001745817	0.0035122589

```
screeplot(pca1,col="cornflowerblue",main = "Scree plot of pca of bigrams
distn in englis.")
```

Scree plot of pca of bigrams
distr in english.



```
#To rank the we sort with respect to first principal components:
ranking=pca1[[2]][order(pca1[[2]][,1],decreasing = TRUE),]
head(ranking[,1],10)
```

```
## those who of the allah is do not of allah allah and and the in the
## 0.2827517 0.2259690 0.1973532 0.1846487 0.1705267 0.1663065 0.1575406 0.1405011
## it is and allah
## 0.1398751 0.1351028
```

```
tail(ranking[,1],10)
```

```
## the deniers i swear righteous will of recompense man will
## -0.0007660194 -0.0007953141 -0.0008023474 -0.0008279591 -0.0008358180
## lord would so which woe that can make you deny
## -0.0011273623 -0.0012364241 -0.0013026014 -0.0015830203 -0.0018261259
```

```
tail1=tail(ranking,20)
ranked=rownames(ranking)

#After ranking we got in top ranked bigrams in English:
head(ranked,100)
```

```
## [1] "those who" "of the" "allah is" "do not"
```

##	[5]	"of allah"	"allah and"	"and the"	"in the"
##	[9]	"it is"	"and allah"	"to you"	"if you"
##	[13]	"indeed allah"	"they will"	"and they"	"is the"
##	[17]	"you and"	"and we"	"to the"	"the people"
##	[21]	"them and"	"will be"	"to them"	"from the"
##	[25]	"allah has"	"for you"	"that which"	"you are"
##	[29]	"and do"	"and if"	"they are"	"and he"
##	[33]	"your lord"	"we have"	"of them"	"you will"
##	[37]	"who have"	"for them"	"the earth"	"what they"
##	[41]	"for the"	"are the"	"there is"	"if they"
##	[45]	"have believed"	"what is"	"to allah"	"o muhammad"
##	[49]	"of what"	"what you"	"will not"	"and when"
##	[53]	"you have"	"that is"	"the scripture"	"is a"
##	[57]	"that you"	"and you"	"and those"	"we will"
##	[61]	"he said"	"but if"	"to him"	"among the"
##	[65]	"and whoever"	"upon you"	"come to"	"lord and"
##	[69]	"and his"	"is no"	"you do"	"he is"
##	[73]	"you who"	"will have"	"that they"	"is not"
##	[77]	"they said"	"on the"	"they were"	"they have"
##	[81]	"who are"	"it and"	"when you"	"for a"
##	[85]	"believe in"	"fear allah"	"for those"	"the day"
##	[89]	"and to"	"o you"	"that allah"	"him and"
##	[93]	"which you"	"does not"	"allah will"	"when they"
##	[97]	"in allah"	"they say"	"the heavens"	"heavens and"

#After ranking we got in last ranked bigrams in English:
`tail(ranked,100)`

##	[1]	"be but"	"for man"	"the crusher"
##	[4]	"avail him"	"be ease"	"hardship will"
##	[7]	"a transaction"	"no you"	"surely see"
##	[10]	"wealth will"	"advised each"	"other to"
##	[13]	"accustomed security"	"saving them"	"created seven"
##	[16]	"in hellfire"	"to deny"	"whoever wills"
##	[19]	"reclining on"	"his provision"	"the hands"
##	[22]	"drink in"	"satisfaction for"	"to feed"
##	[25]	"you into"	"inevitable reality"	"the inevitable"
##	[28]	"had thought"	"the later"	"noble messenger"
##	[31]	"seized him"	"reward uninterrupted"	"mixture is"
##	[34]	"proceed to"	"i invited"	"what are"
##	[37]	"his record"	"favor to"	"another to"
##	[40]	"to drink"	"a cup"	"cup of"
##	[43]	"how it"	"so remind"	"recorded in"
##	[46]	"reached you"	"gardens and"	"the spirit"
##	[49]	"and proportioned"	"burn therein"	"did he"
##	[52]	"heaven is"	"exalt the"	"qur'an easy"
##	[55]	"of certainty"	"the warning"	"eyes humbled"
##	[58]	"created man"	"it covers"	"will ease"
##	[61]	"for remembrance"	"remembrance so"	"severe were"
##	[64]	"were my"	"night of"	"of decree"
##	[67]	"raised high"	"there reached"	"wills may"
##	[70]	"no but"	"out the"	"how severe"

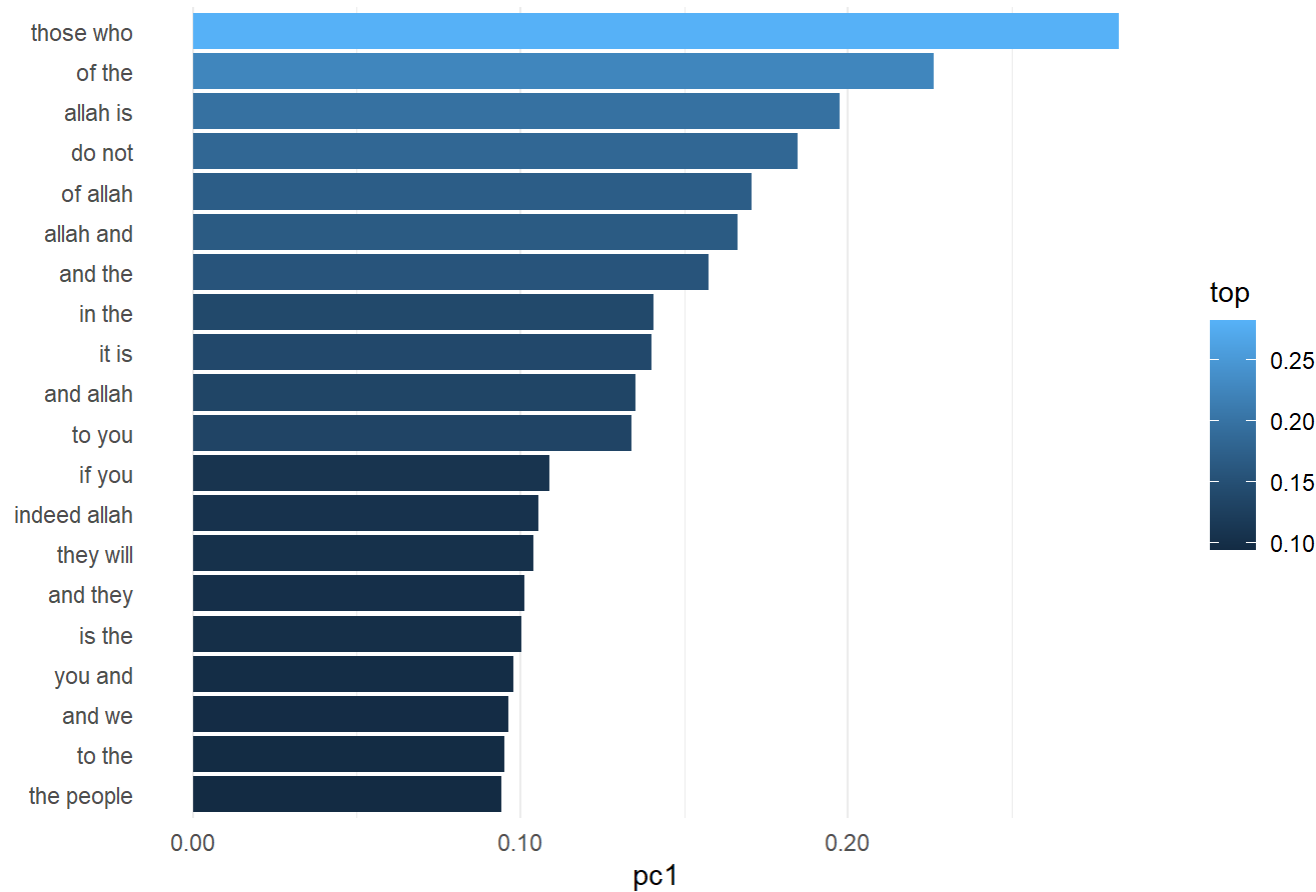
##	[73]	"see it"	"striking calamity"	"the striking"
##	[76]	"a madman"	"mountains are"	"mountains will"
##	[79]	"easy for"	"day will"	"you seen"
##	[82]	"night when"	"burn in"	"enter to"
##	[85]	"to burn"	"lord a"	"there any"
##	[88]	"no indeed"	"faces that"	"and by"
##	[91]	"the deniers"	"i swear"	"righteous will"
##	[94]	"of recompense"	"man will"	"lord would"
##	[97]	"so which"	"woe that"	"can make"
##	[100]	"you deny"		

#To plot Top 20 words in the ranking

```
ranking1=data.frame(x1=rownames(ranking)[1:20], top=ranking[1:20,1])

ggplot(ranking1, aes(x = fct_rev(reorder(x1,-top)),
                    y = top, fill = top)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Top 20 bigrams in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

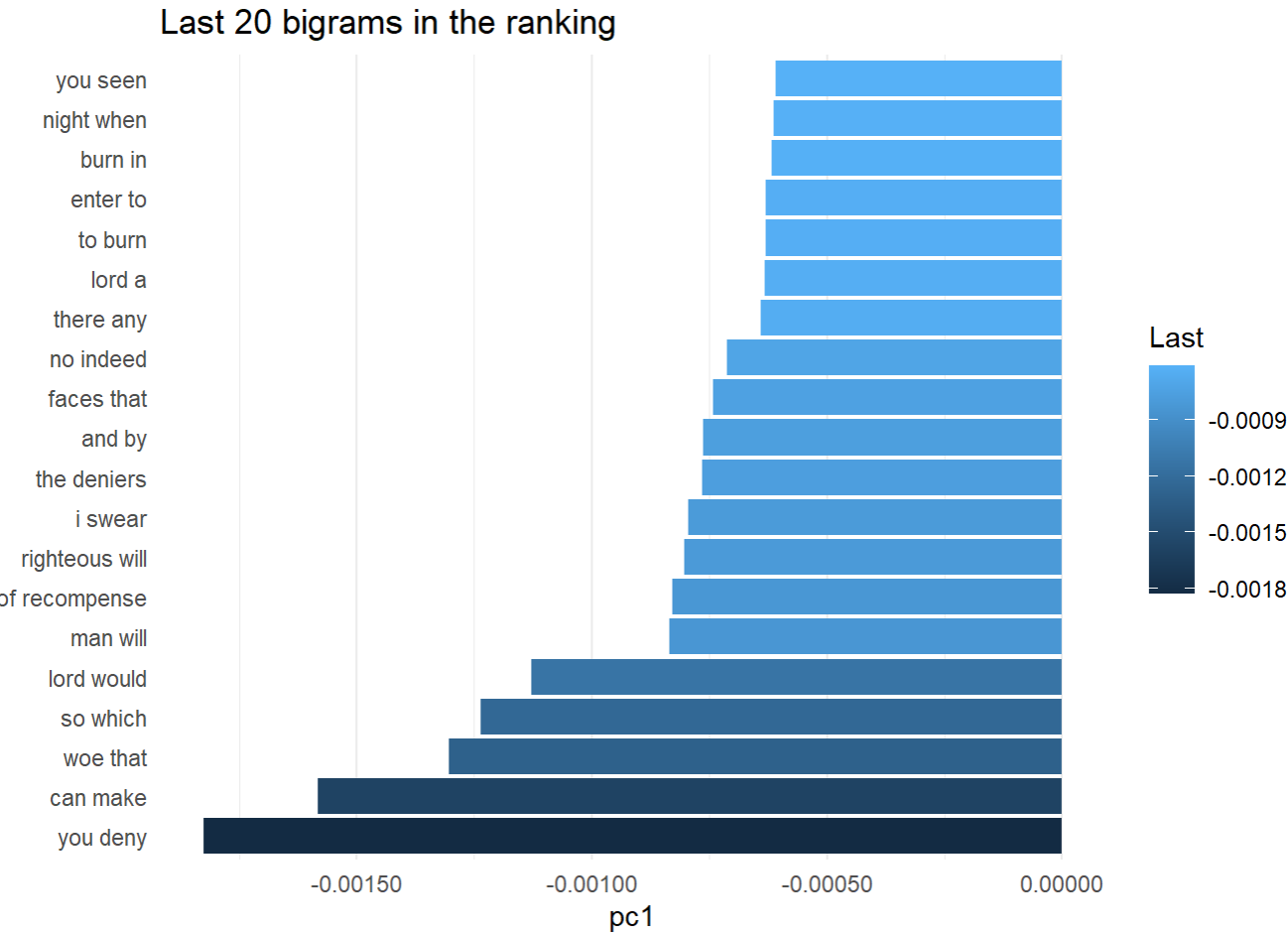
Top 20 bigrams in the ranking



#To plot Last 20 words in the ranking


```
ranking2=data.frame(x1=rownames(tail1),Last=tail1[,1])

ggplot(ranking2, aes(x = fct_rev(reorder(x1,-Last)),
                    y = Last, fill = Last)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Last 20 bigrams in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
dbigram1=as.data.frame(dbigram)

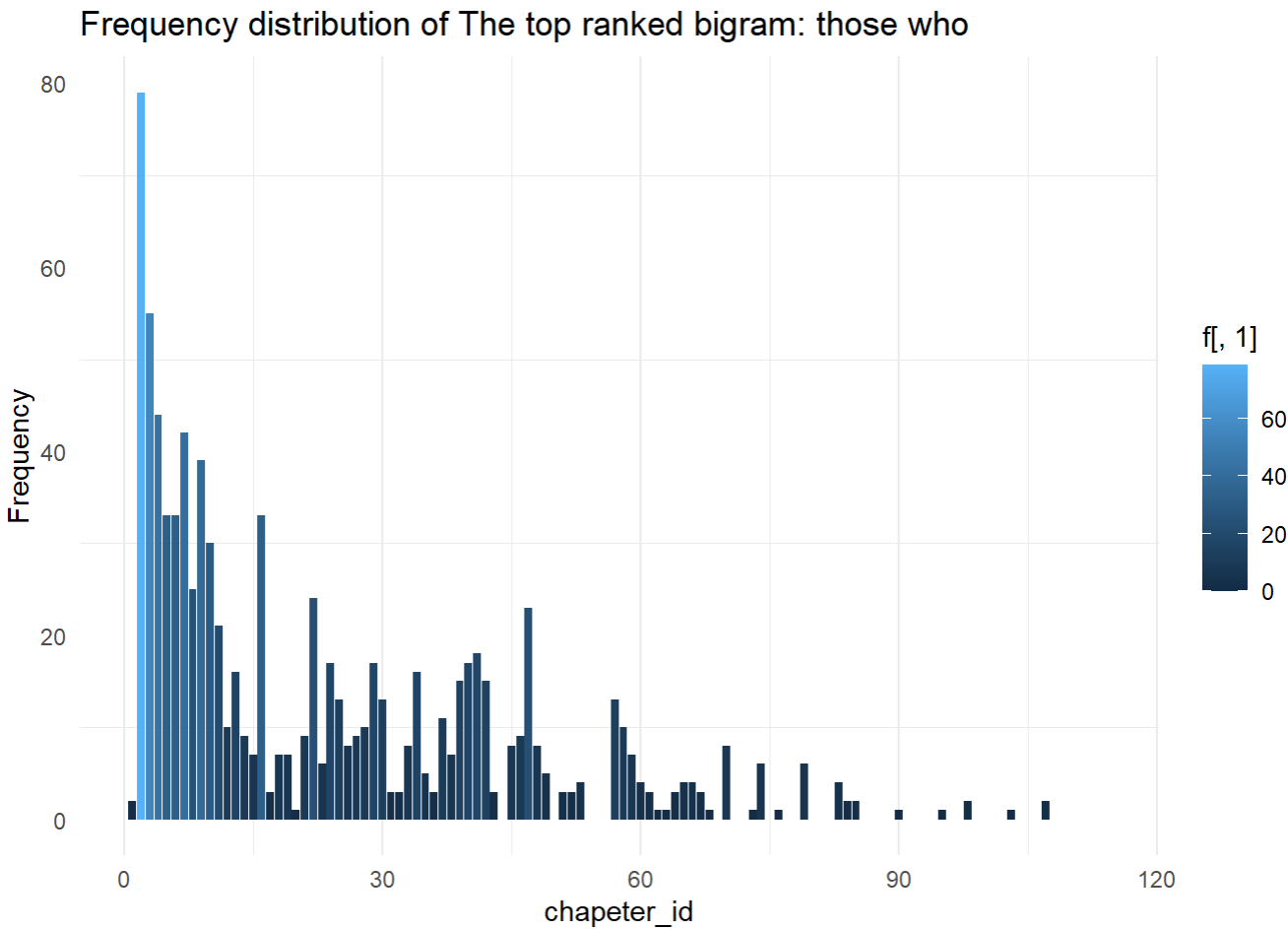
f=as.matrix(filter(dbigram1 ,rownames(dbigram1) %in% head(ranked,1)))
f1=as.matrix(filter(dbigram1 ,rownames(dbigram1) %in% tail(ranked,1)))

#view(f)
#view(f1)

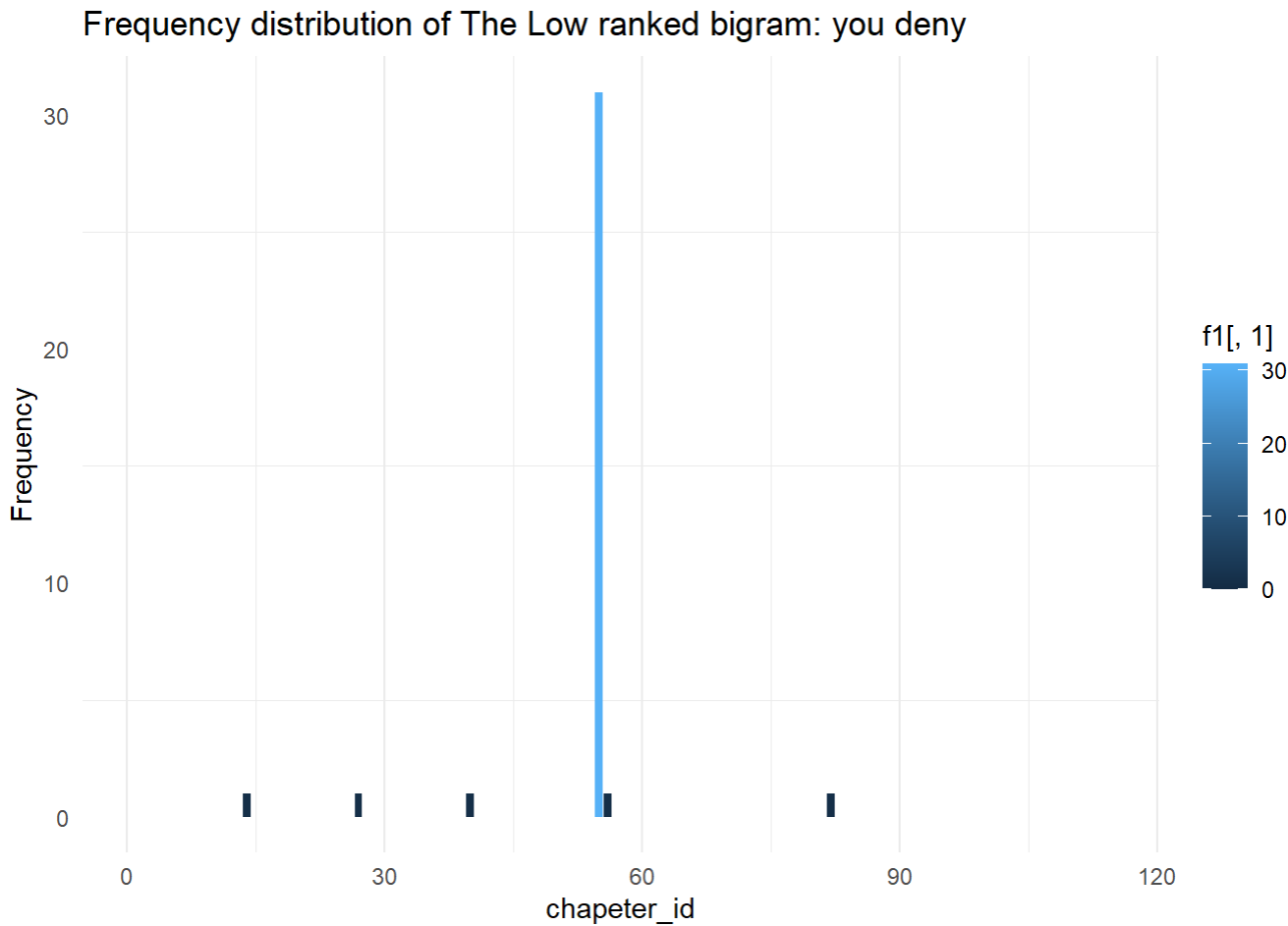
f=as.data.frame(t(f))
f=cbind(f,chapter_id=1:114)

f1=as.data.frame(t(f1))
f1=cbind(f1,chapter_id=1:114)
```

```
#Frequency distribution of The top ranked word
ggplot(f,aes(fill=f[,1],x=chapter_id,y=f[,1]))+
geom_col()+
labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The top ranked bigram: those who") +
theme_minimal() +
theme(panel.grid.major.y = element_blank())
```



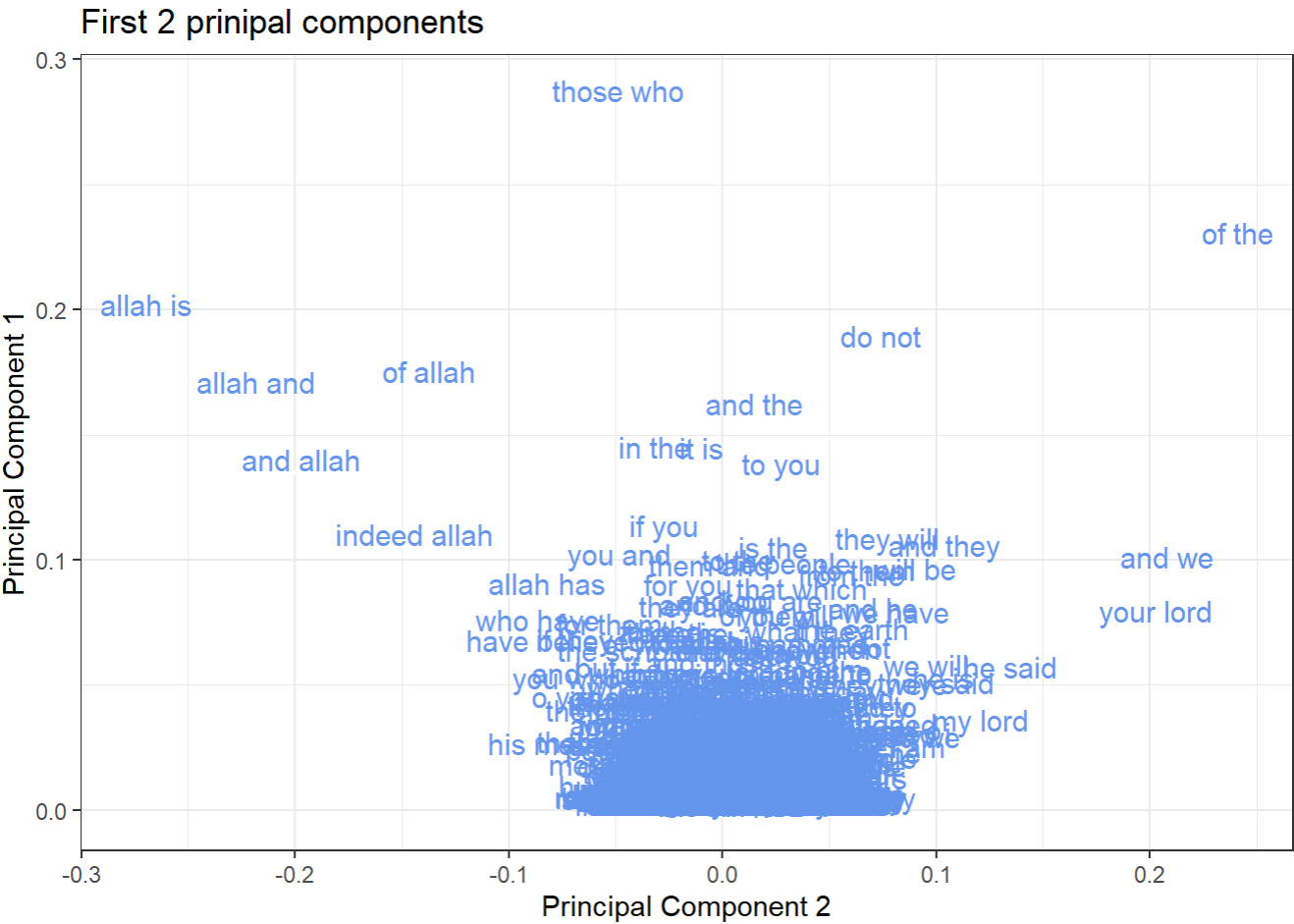
```
ggplot(f1,aes(fill=f1[,1],x=chapter_id,y=f1[,1]))+
geom_col()+
labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The Low ranked bigram: you deny") +
theme_minimal() +
theme(panel.grid.major.y = element_blank())
```



*#To Detect multidimensional outliers, if any, from the projection of
#the data:*

```
pc12=data.frame(pca1$rotation[,c(1,2)])

ggplot(pc12,
  aes(x=PC2,
    y=PC1)) +
  geom_point(color="white") +
  theme_bw() +
  ggtitle('First 2 prinipal components') +
  xlab("Principal Component 2") +
  ylab("Principal Component 1")+
  geom_text(label=rownames(pc12),
    nudge_x = 0.005,
    nudge_y = 0.005,
    size=4,colour="cornflowerblue")
```



#To ranke of the Holy Qur'an trigrams in arabic:
#####

```
book_id=rep(1,nrow(quran_ar_min))
quran_ar_min1=cbind(book_id,quran_ar_min)
head(quran_ar_min1)
```

##	book_id	surah_id	ayah_id	surah_title_ar	surah_title_en	surah_title_en_trans		
## 1	1	1	1	الفاتحة	Al-Faatiha	The Opening		
## 2	1	1	2	الفاتحة	Al-Faatiha	The Opening		
## 3	1	1	3	الفاتحة	Al-Faatiha	The Opening		
## 4	1	1	4	الفاتحة	Al-Faatiha	The Opening		
## 5	1	1	5	الفاتحة	Al-Faatiha	The Opening		
## 6	1	1	6	الفاتحة	Al-Faatiha	The Opening		
##	revelation_type	text surah ayah			ayah_title	juz		
## 1	Meccan	<U+FEFF>1	1:1	1	1	بسم الله الرحمن الرحيم		
## 2	Meccan		1 1:2	2	1	الحمد لله رب العالمين		
## 3	Meccan		1 1:3	3	1	الرحمن الرحيم		
## 4	Meccan		1 1:4	4	1	مالك يوم الدين		
## 5	Meccan	1	1:5	5	1	إياك نعبد وإياك نستعين		
## 6	Meccan	1	1:6	6	1	اهدنا الصراط المستقيم		
##	manzil	page	hizb	quarter	sajda	sajda_id	sajda_recommended	sajda_obligatory
## 1	1	1	1	FALSE	NA	NA	NA	NA
## 2	1	1	1	FALSE	NA	NA	NA	NA
## 3	1	1	1	FALSE	NA	NA	NA	NA

## 4	1	1	1 FALSE	NA	NA	NA
## 5	1	1	1 FALSE	NA	NA	NA
## 6	1	1	1 FALSE	NA	NA	NA

```
quran_words <- quran_ar_min1 %>%
  unnest_tokens(trigram, text, token = "ngrams", n = 3) %>%
  count(book_id, trigram, sort = TRUE)

colnames(quran_words)=c("book_id","trigram","freq_in_book")

head(quran_words)
```

##	book_id	trigram	freq_in_book
## 1	1	<NA>	163
## 2	1 114	الله الرحمن الرحيم	
## 3	1 114	بسم الله الرحمن	
## 4	1 92	يا أيها الذين	
## 5	1 89	أيها الذين آمنوا	
## 6	1 71	من دون الله	

```
chapter_words <- quran_ar_min1 %>%
  unnest_tokens(trigram, text, token = "ngrams", n = 3) %>%
  count(surah_title_en, trigram, sort = TRUE)

colnames(chapter_words)=c("surah_title_en","trigram","freq_in_chapter")

chapter_words=na.omit(chapter_words)
head(chapter_words)
```

##	surah_title_en	trigram	freq_in_chapter
## 1	Ar-Rahmaan 31	آلاء ربكما تكذبان	
## 2	Ar-Rahmaan 31	فبأي آلاء ربكما	
## 3	An-Nisaa 16	إن الله كان	
## 4	Al-Maaida 16	أيها الذين آمنوا	
## 5	Al-Maaida 16	يا أيها الذين	
## 9	Al-Baqara 11	أيها الذين آمنوا	

```
quran_words=left_join(quran_words,chapter_words)
```

```
## Joining, by = "trigram"
```

```
#arabic_stopwords = data.frame(word = removeStopWords("ÓäÇä")$arabicStopwordList)

#book_words=anti_join(book_words,arabic_stopwords, by = "word")
head(quran_words)
```

##	book_id	trigram	freq_in_book	surah_title_en	freq_in_chapter
## 1	1	<NA>	163	<NA>	NA

## 2	1 114	الله الرحمن الرحيم	An-Naml	2
## 3	1 114	الله الرحمن الرحيم	Al-Faatiha	1
## 4	1 114	الله الرحمن الرحيم	Al-Baqara	1
## 5	1 114	الله الرحمن الرحيم	Aal-i-Imraan	1
## 6	1 114	الله الرحمن الرحيم	An-Nisaa	1

```
token_list=as.data.frame(table(chapter_words$surah_title_en,chapter_words$strigram))
token_list=token_list[,-3]
colnames(token_list)=c("surah_title_en","trigram")
head(token_list)
```

```
## surah_title_en trigram
## 1 Al-Faatiha ن والقلم وما
## 2 Al-Baqara ن والقلم وما
## 3 Aal-i-Imraan ن والقلم وما
## 4 An-Nisaa ن والقلم وما
## 5 Al-Maaida ن والقلم وما
## 6 Al-An'aam ن والقلم وما
```

```
#book_words=mutate(book_words,ratio=book_words$freq_in_chapter/
# book_words$freq_in_book)
head(quran_words)
```

##	book_id	trigram	freq_in_book	surah_title_en	freq_in_chapter
## 1	1	<NA>	163	<NA>	NA
## 2	1 114	الله الرحمن الرحيم		An-Naml	2
## 3	1 114	الله الرحمن الرحيم		Al-Faatiha	1
## 4	1 114	الله الرحمن الرحيم		Al-Baqara	1
## 5	1 114	الله الرحمن الرحيم		Aal-i-Imraan	1
## 6	1 114	الله الرحمن الرحيم		An-Nisaa	1

```
quran_words=left_join(token_list,quran_words,by=c("surah_title_en","trigram"))
quran_words[is.na(quran_words)]=0
head(quran_words)
```

##	surah_title_en	trigram	book_id	freq_in_book	freq_in_chapter
## 1	Al-Faatiha	0	0	0	ن والقلم وما
## 2	Al-Baqara	0	0	0	ن والقلم وما
## 3	Aal-i-Imraan	0	0	0	ن والقلم وما
## 4	An-Nisaa	0	0	0	ن والقلم وما
## 5	Al-Maaida	0	0	0	ن والقلم وما
## 6	Al-An'aam	0	0	0	ن والقلم وما

```
r=unique(quran_words$strigram)
c=unique(quran_ar_min$surah_title_en)

dtrigram=matrix(quran_words$freq_in_chapter,nrow=nrow(quran_words)/114,
ncol=114,byrow=T,dimnames=list(r,c))
```

head(dtrigram)

##	Al-Faatiha Al-Baqara Aal-i-Imraan An-Nisaa Al-Maaida Al-An'aam						
## 0	0	0	0	0	0		ن والقلم وما
## 0	0	0	0	0	0		إن هم إلا
## 0	0	0	0	0	0		له إن هو
## 0	0	0	0	0	0		له من نور
## 0	0	0	0	0	0		له من أمره
## 0	0	0	0	0	0		له من أمرنا
##	Al-A'raaf Al-Anfaal At-Tawba Yunus Hud Yusuf Ar-Ra'd Ibrahim						
## 0	0	0	0	0	0	0	ن والقلم وما
## 0	0	0	0	0	0	0	إن هم إلا
## 0	0	0	0	0	0	0	له إن هو
## 0	0	0	0	0	0	0	له من نور
## 0	0	0	0	0	0	0	له من أمره
## 0	0	0	0	0	0	0	له من أمرنا
##	Al-Hijr An-Nahl Al-Israa Al-Kahf Maryam Taa-Haa Al-Anbiyaa Al-Hajj						
## 0	0	0	0	0	0	0	ن والقلم وما
## 0	0	0	0	0	0	0	إن هم إلا
## 0	0	0	0	0	0	0	له إن هو
## 0	0	0	0	0	0	0	له من نور
## 0	0	0	0	0	0	0	له من أمره
## 0	0	0	0	1	0	0	له من أمرنا
##	Al-Muminoon An-Noor Al-Furqaan Ash-Shu'araa An-Naml Al-Qasas						
## 0	0	0	0	0	0		ن والقلم وما
## 0	0	0	1	0	0		إن هم إلا
## 0	0	0	0	0	0		له إن هو
## 0	0	0	0	1	0		له من نور
## 0	0	0	0	0	0		له من أمره
## 0	0	0	0	0	0		له من أمرنا
##	Al-Ankaboot Ar-Room Luqman As-Sajda Al-Ahzaab Saba Faatir Yaseen						
## 0	0	0	0	0	0	0	ن والقلم وما
## 0	0	0	0	0	0	0	إن هم إلا
## 1	0	0	0	0	0	0	له إن هو
## 0	0	0	0	0	0	0	له من نور
## 0	0	0	0	0	0	0	له من أمره
## 0	0	0	0	0	0	0	له من أمرنا
##	As-Saaffaat Saad Az-Zumar Ghafir Fussilat Ash-Shura Az-Zukhruf						
## 0	0	0	0	0	0	0	ن والقلم وما
## 1	0	0	0	0	0	0	إن هم إلا
## 0	0	0	0	0	0	0	له إن هو
## 0	0	0	0	0	0	0	له من نور
## 0	0	0	0	0	0	0	له من أمره
## 0	0	0	0	0	0	0	له من أمرنا
##	Ad-Dukhaan Al-Jaathiya Al-Ahqaf Muhammad Al-Fath Al-Hujuraat Qaaf						
## 0	0	0	0	0	0	0	ن والقلم وما
## 0	0	0	0	0	1	0	إن هم إلا
## 0	0	0	0	0	0	0	له إن هو
## 0	0	0	0	0	0	0	له من نور
## 0	0	0	0	0	0	0	له من أمره
## 0	0	0	0	0	0	0	له من أمرنا

##	0	0	0	0	0	0	
##	Adh-Dhaariyat At-Tur An-Najm Al-Qamar Ar-Rahmaan Al-Waaqia						
##	0	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو
##	0	0	0	0	0	0	له من نور
##	0	0	0	0	0	0	له من أمره
##	0	0	0	0	0	0	له من أمرنا
##	Al-Hadid Al-Mujaadila Al-Hashr Al-Mumtahana As-Saff Al-Jumu'a						
##	0	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو
##	0	0	0	0	0	0	له من نور
##	0	0	0	0	0	0	له من أمره
##	0	0	0	0	0	0	له من أمرنا
##	Al-Munaafiqoon At-Taghaabun At-Talaaq At-Tahrim Al-Mulk Al-Qalam						
##	1	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو
##	0	0	0	0	0	0	له من نور
##	0	0	1	0	0	0	له من أمره
##	0	0	0	0	0	0	له من أمرنا
##	Al-Haaqqa Al-Ma'aarij Nooh Al-Jinn Al-Muzzammil Al-Muddaththir						
##	0	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو
##	0	0	0	0	0	0	له من نور
##	0	0	0	0	0	0	له من أمره
##	0	0	0	0	0	0	له من أمرنا
##	Al-Qiyaama Al-Insaan Al-Mursalaat An-Naba An-Naazi'aat Abasa						
##	0	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو
##	0	0	0	0	0	0	له من نور
##	0	0	0	0	0	0	له من أمره
##	0	0	0	0	0	0	له من أمرنا
##	At-Takwir Al-Infitaar Al-Mutaffifin Al-Inshiqaaq Al-Burooj						
##	0	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو
##	0	0	0	0	0	0	له من نور
##	0	0	0	0	0	0	له من أمره
##	0	0	0	0	0	0	له من أمرنا
##	At-Taariq Al-A'laa Al-Ghaashiya Al-Fajr Al-Balad Ash-Shams Al-Lail						
##	0	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو
##	0	0	0	0	0	0	له من نور
##	0	0	0	0	0	0	له من أمره
##	0	0	0	0	0	0	له من أمرنا
##	Ad-Dhuhaa Ash-Sharh At-Tin Al-Alaq Al-Qadr Al-Bayyina Az-Zalzala						
##	0	0	0	0	0	0	ن والقلم وما
##	0	0	0	0	0	0	إن هم إلا
##	0	0	0	0	0	0	له إن هو


```
## 0 0 0 0 0 0 0 له من نور
## 0 0 0 0 0 0 0 له من أمره
## 0 0 0 0 0 0 0 له من أمرنا
## Al-Aadiyaat Al-Qaari'a At-Takaathur Al-Asr Al-Humaza Al-Fil
## 0 0 0 0 0 0 ن والقلم وما
## 0 0 0 0 0 0 إن هم إلا
## 0 0 0 0 0 0 له إن هو
## 0 0 0 0 0 0 له من نور
## 0 0 0 0 0 0 له من أمره
## 0 0 0 0 0 0 له من أمرنا
## Quraish Al-Maa'un Al-Kawthar Al-Kaafiroon An-Nasr Al-Masad
## 0 0 0 0 0 0 ن والقلم وما
## 0 0 0 0 0 0 إن هم إلا
## 0 0 0 0 0 0 له إن هو
## 0 0 0 0 0 0 له من نور
## 0 0 0 0 0 0 له من أمره
## 0 0 0 0 0 0 له من أمرنا
## Al-Ikhlaas Al-Falaq An-Naas
## 0 0 0 ن والقلم وما
## 0 0 0 إن هم إلا
## 0 0 0 له إن هو
## 0 0 0 له من نور
## 0 0 0 له من أمره
## 0 0 0 له من أمرنا
```

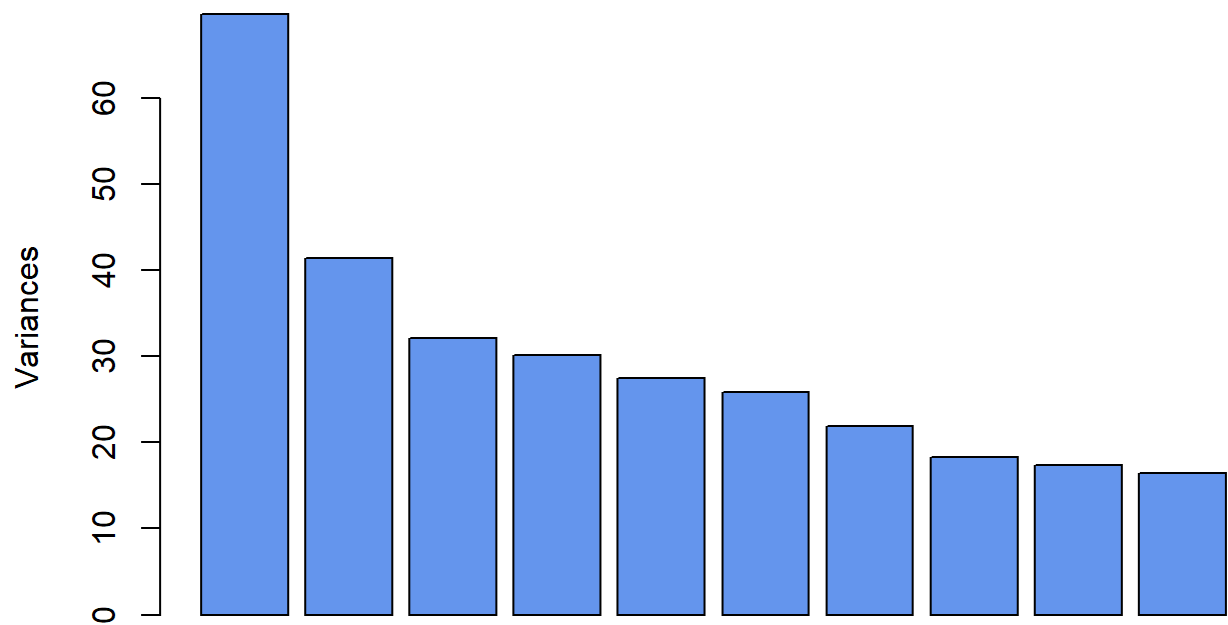
```
pca1=prcomp(t(dtrigram))

pc12=data.frame(pca1$rotation[,c(1,2)])
head(pc12)
```

```
##          PC1          PC2
## 0.0002129668- 0.0002550257 ن والقلم وما
## 0.0005569569- 0.0005898122 إن هم إلا
## 0.0002665095- 0.0002207569 له إن هو
## 0.0003470954 0.0003149021- له من نور
## 0.0001103400- 0.0001560399 له من أمره
## 0.0002816484- 0.0001999607 له من أمرنا
```

```
screepLOT(pca1,col="cornflowerblue",main = "Scree plot of pca of trigrams
distr in arabic.")
```

Scree plot of pca of trigrams distn in arabic.



#To rank the we sort with respect to first principal components:

```
ranking=pca1[[2]][order(pca1[[2]][,1],decreasing = TRUE),]  
head(ranking[,1],10)
```

```
## آلاء ربكما تكذبان فبأي آلاء ربكما وما أدراك ما ويل يومئذ للمكذبين  
## 0.010545238 0.010545238 0.003051940 0.002800913  
## في ذلك لآيات رب السماوات والأرض مع الله إلهها الله إلهها آخر  
## 0.002737664 0.002502415 0.002167363 0.002167363  
## وقال الذين كفروا لهو العزيز الرحيم  
## 0.001998849 0.001982147
```

```
tail(ranking[,1],10)
```

```
## هم فيها خالدون الله غفور رحيم بالله واليوم الآخر كل شيء قدير  
## -0.06588633 -0.06708387 -0.07894250 -0.08081398  
## واعلموا أن الله على كل شيء من بعد ما في سبيل الله  
## -0.08439477 -0.09058098 -0.09971178 -0.14894323  
## أيها الذين آمنوا يا أيها الذين  
## -0.20175954 -0.20405359
```

```
tail1=tail(ranking,20)  
ranked=rownames(ranking)
```

#After ranking we got in top ranked trigrams in Arabic:

head(ranked,100)

"آلاء ربكما تكذبان" "فبأي آلاء ربكما" "وما أدراك ما" [1] ##
"ويل يومئذ للمكذبين" "في ذلك لآيات" "رب السماوات والأرض" [4] ##
"مع الله إلهها" "الله إلهها آخر" "وقال الذين كفروا" [7] ##
"لهو العزيز الرحيم" "لآية وما كان" "وإن ربك لهو" [10] ##
"وما كان أكثرهم" "ذلك لآية وما" "ربك لهو العزيز" [13] ##
"كان أكثرهم مؤمنين" "عليه من أجر" "في ضلال مبين" [16] ##
"إن هو إلا" "أسألكم عليه من" "الرحمن الرحيم إذا" [19] ##
"فهل من مدكر" "وهم لا يشعرون" "كانوا به يستهزئون" [22] ##
"إن المتقين في" "كذلك نجزي المحسنين" "إن أجري إلا" [25] ##
"أجري إلا على" "إني لكم رسول" "لكم رسول أمين" [28] ##
"من أجر إن" "الناس لا يعلمون" "الله الرحمن الرحيم" [31] ##
"بسم الله الرحمن" "ما كنتم تعملون" "رحمة من ربك" [34] ##
"الحمد لله الذي" "الرحمن الرحيم قل" "إنا كذلك نجزي" [37] ##
"عباد الله المخلصين" "وعد الله حق" "الرزق لمن يشاء" [40] ##
"يبسط الرزق لمن" "إلا على رب" "إذ قال لهم" [43] ##
"وما أسألكم عليه" "على رب العالمين" "متنا وكنا ترابا" [46] ##
"إلا الذين آمنوا" "الذي كنتم به" "ويقولون متى هذا" [49] ##
"هو إلا ذكر" "ذلك لآيات لقوم" "لا يؤمنون بالآخرة" [52] ##
"أجر إن أجري" "قالوا يا أبانا" "الرحمن الرحيم حم" [55] ##
"ما كانوا به" "كان عاقبة الذين" "ما هذا إلا" [58] ##
"واذكر في الكتاب" "أهلكنا قبلهم من" "أإله مع الله" [61] ##
"للذكر فهل من" "ولقد يسرنا القرآن" "القرآن للذكر فهل" [64] ##
"يسرنا القرآن للذكر" "كان عذابي ونذر" "فكيف كان عذابي" [67] ##
"في هذا القرآن" "أرسلنا من قبلك" "المتقين في جنات" [70] ##
"بما كانوا يعملون" "فإذا جاء وعد" "يا أيها الملأ" [73] ##
"ولم يكن له" "من أوتي كتابه" "من عبادنا المؤمنين" [76] ##
"إلا عباد الله" "في جنات وعيون" "كنت من الصادقين" [79] ##
"إن أنتم إلا" "تلك آيات الكتاب" "قال لهم أخوهم" [82] ##
"من قبلهم من" "هذا الوعد إن" "متى هذا الوعد" [85] ##
"الوعد إن كنتم" "كنتم به تكذبون" "أإذا متنا وكنا" [88] ##
"إنه من عبادنا" "هل أتاك حديث" "وكنا ترابا وعظاما" [91] ##
"ترابا وعظاما أينا" "إلا ذكر للعالمين" "قل أرأيتم إن" [94] ##
"أن يشاء الله" "بل هم في" "إذ قال لأبيه" [97] ##
"قال يا قوم" [100]

#After ranking we got in last ranked trigrams in Arabic:

tail(ranked,100)

"وإذ أخذنا ميثاقكم" "وجهك شطر المسجد" [1] ##
"إسرائيل اذكروا نعمتي" "نعمتي التي أنعمت" [3] ##
"التي أنعمت عليكم" "اذكروا نعمتي التي" [5] ##
"بني إسرائيل اذكروا" "تنفقوا من خير" [7] ##
"بالمعروف حقا على" "يبين لنا ما" [9] ##
"حدود الله فلا" "يقول إنها بقرة" [11] ##
"يقيما حدود الله" "ربك يبين لنا" [13] ##
"شطر المسجد الحرام" "قال إنه يقول" [15] ##
"قالوا ادع لنا" "قول وجهك شطر" [17] ##
"يهدي من يشاء" "والله لا يهدي" [19] ##
"في الدنيا والآخرة" "ولهم عذاب أليم" [21]

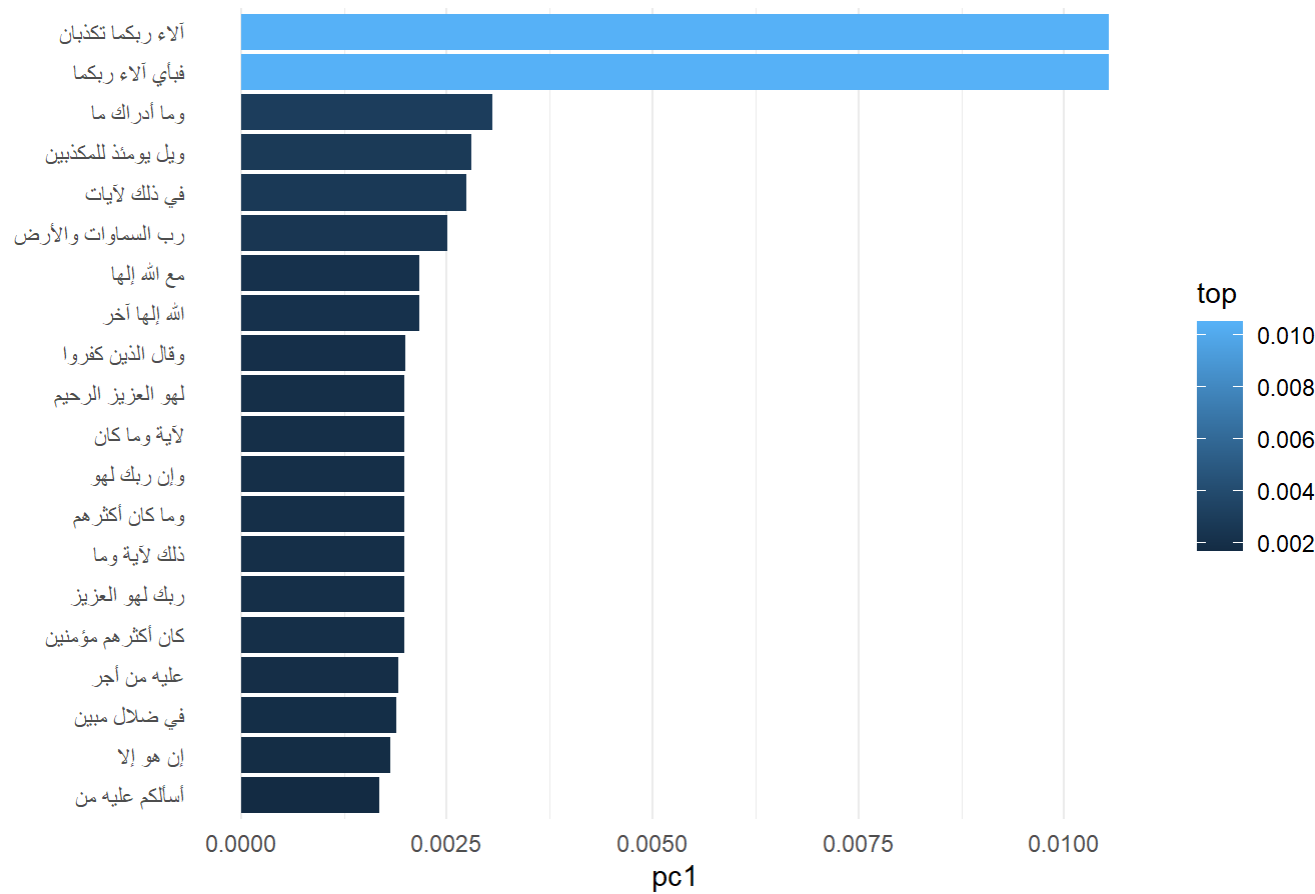
[23] "يؤمن بالله واليوم" "الله سميع عليم"
[25] "لا إله إلا" "من أهل الكتاب"
[27] "الله عزيز حكيم" "إله إلا هو"
[29] "يا بني إسرائيل" "فلا جناح عليهما"
[31] "فلا جناح عليكم" "الذين آمنوا والذين"
[33] "وأقيموا الصلاة وآتوا" "إن الله على"
[35] "وما تنفقوا من" "جنان تجري من"
[37] "لكم إن كنتم" "بما أنزل الله"
[39] "إن الذين آمنوا" "لا يهدي القوم"
[41] "إبراهيم وإسماعيل وإسحاق" "والله واسع عليم"
[43] "عيسى ابن مريم" "بكل شيء عليم"
[45] "يبين الله لكم" "من بعد ذلك"
[47] "إن الله غفور" "الذين ينفقون أموالهم"
[49] "ربهم ولا خوف" "عند ربهم ولا"
[51] "فلا إثم عليه" "إن الله كان"
[53] "ومن الناس من" "آمنوا وعملوا الصالحات"
[55] "وما في الأرض" "السموات وما في"
[57] "في السموات وما" "إن الذين كفروا"
[59] "من يشاء والله" "بغافل عما تعملون"
[61] "بما تعملون بصير" "أجرهم عند ربهم"
[63] "وإذا قيل لهم" "ألم تر إلى"
[65] "إن الله لا" "من تحتها الأنهار"
[67] "تجري من تحتها" "ولو شاء الله"
[69] "إن كنتم صادقين" "الصلاة وآتوا الزكاة"
[71] "كذلك يبين الله" "والله بما تعملون"
[73] "من دون الله" "ولا خوف عليهم"
[75] "الله واعلموا أن" "الذين أوتوا الكتاب"
[77] "وما الله بغافل" "الله بغافل عما"
[79] "أصحاب النار هم" "النار هم فيها"
[81] "الذين آمنوا لا" "ما في السموات"
[83] "الله على كل" "إن كنتم مؤمنين"
[85] "من عند الله" "واتقوا الله واعلموا"
[87] "الله إن الله" "ولا هم يحزنون"
[89] "خوف عليهم ولا" "عليهم ولا هم"
[91] "هم فيها خالدون" "الله غفور رحيم"
[93] "بالله واليوم الآخر" "كل شيء قدير"
[95] "واعلموا أن الله" "على كل شيء"
[97] "من بعد ما" "في سبيل الله"
[99] "أيها الذين آمنوا" "يا أيها الذين"

#To plot Top 20 words in the ranking

```
ranking1=data.frame(x1=rownames(ranking)[1:20], top=ranking[1:20,1])

ggplot(ranking1, aes(x = fct_rev(reorder(x1,-top)),
  y = top, fill = top)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Top 20 trigrams in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

Top 20 trigrams in the ranking

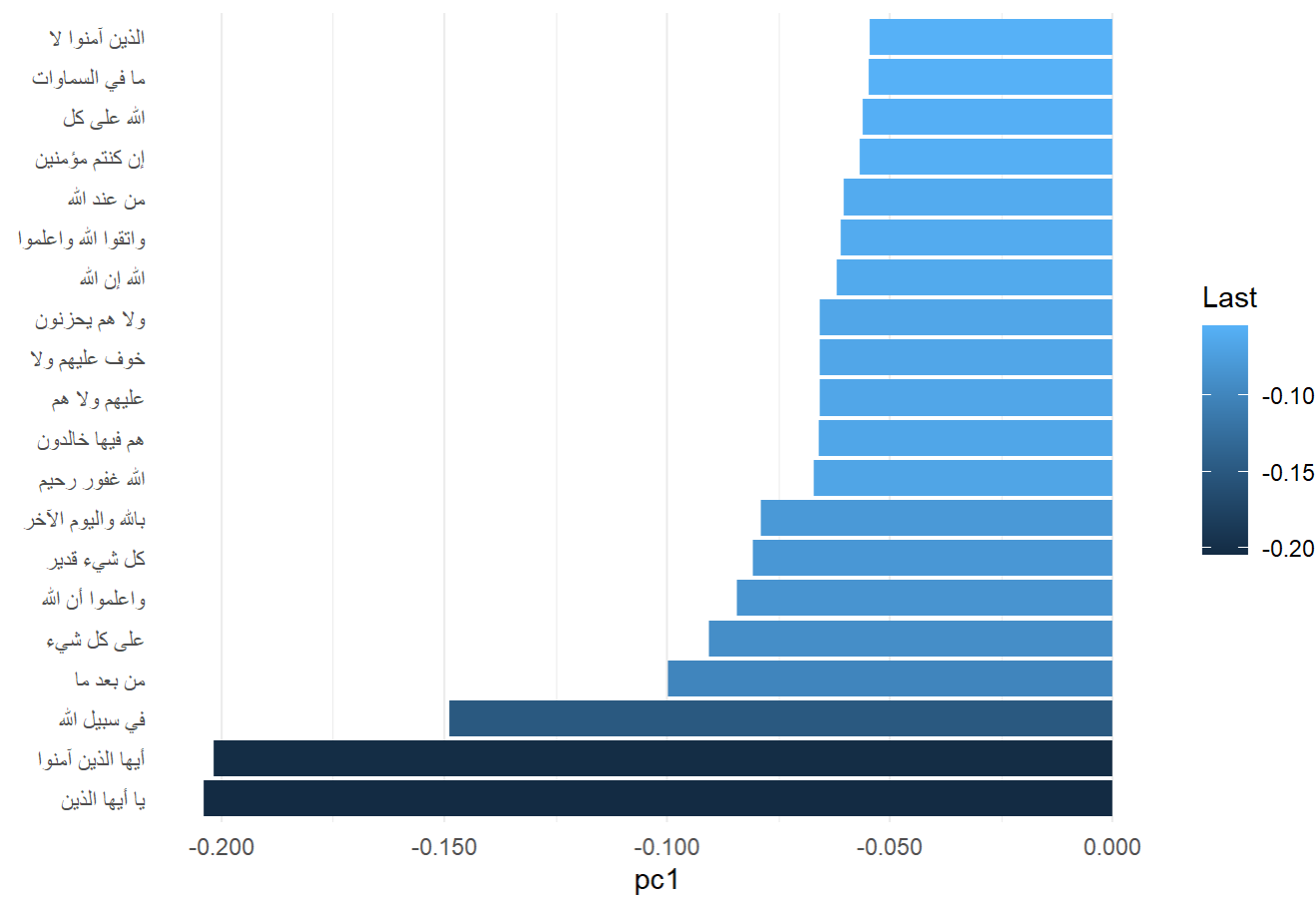


#To plot Last 20 words in the ranking

```
ranking2=data.frame(x1=rownames(tail1),Last=tail1[,1])

ggplot(ranking2, aes(x = fct_rev(reorder(x1,-Last)),
  y = Last, fill = Last)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Last 20 trigrams in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

Last 20 trigrams in the ranking



```
dtrigram=as.data.frame(dtrigram)

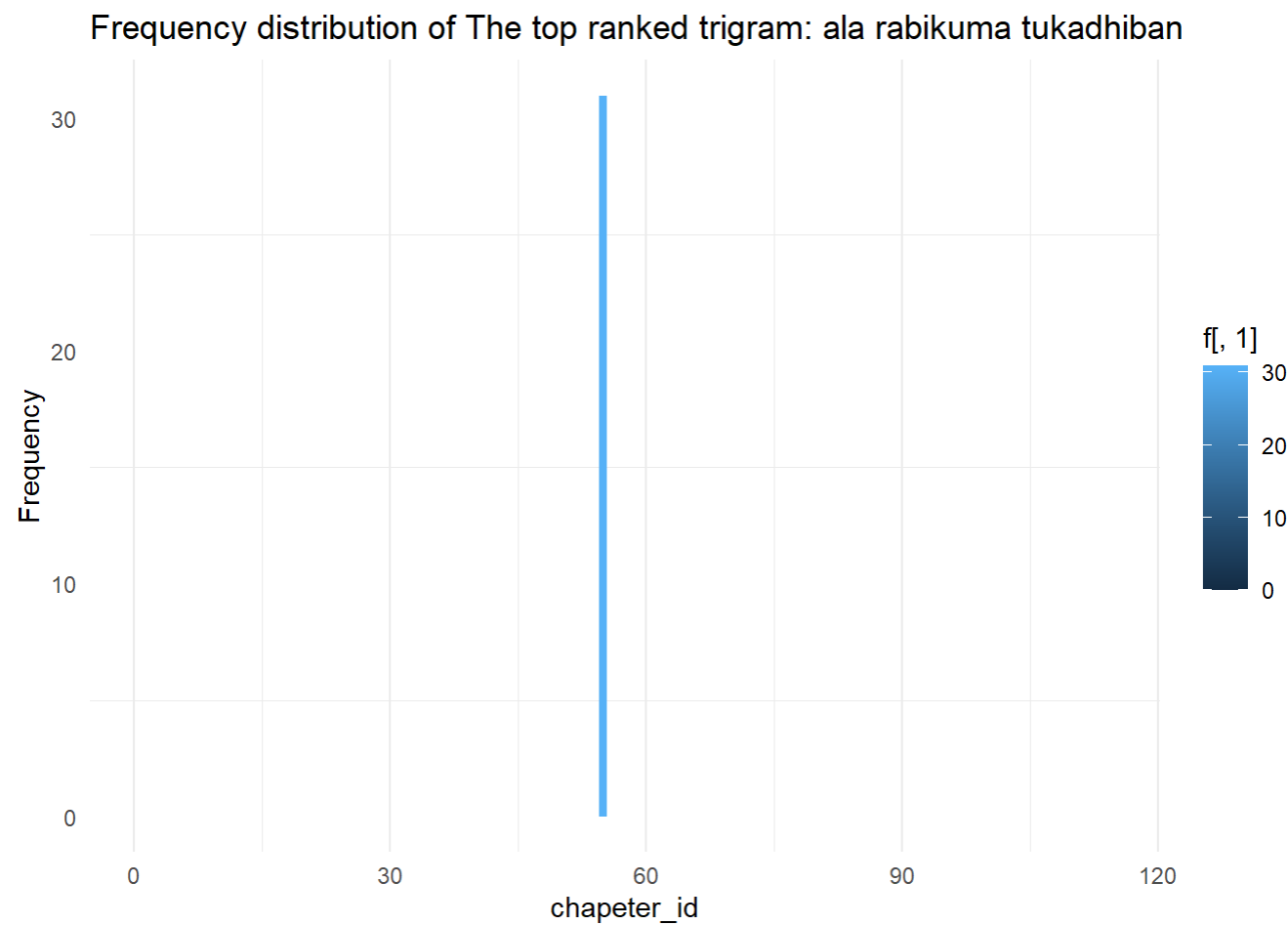
f=filter(dtrigram ,rownames(dtrigram) %in% head(ranked,1))
fl=filter(dtrigram ,rownames(dtrigram) %in% tail(ranked,1))

#view(t(f))
#view(t(fl))

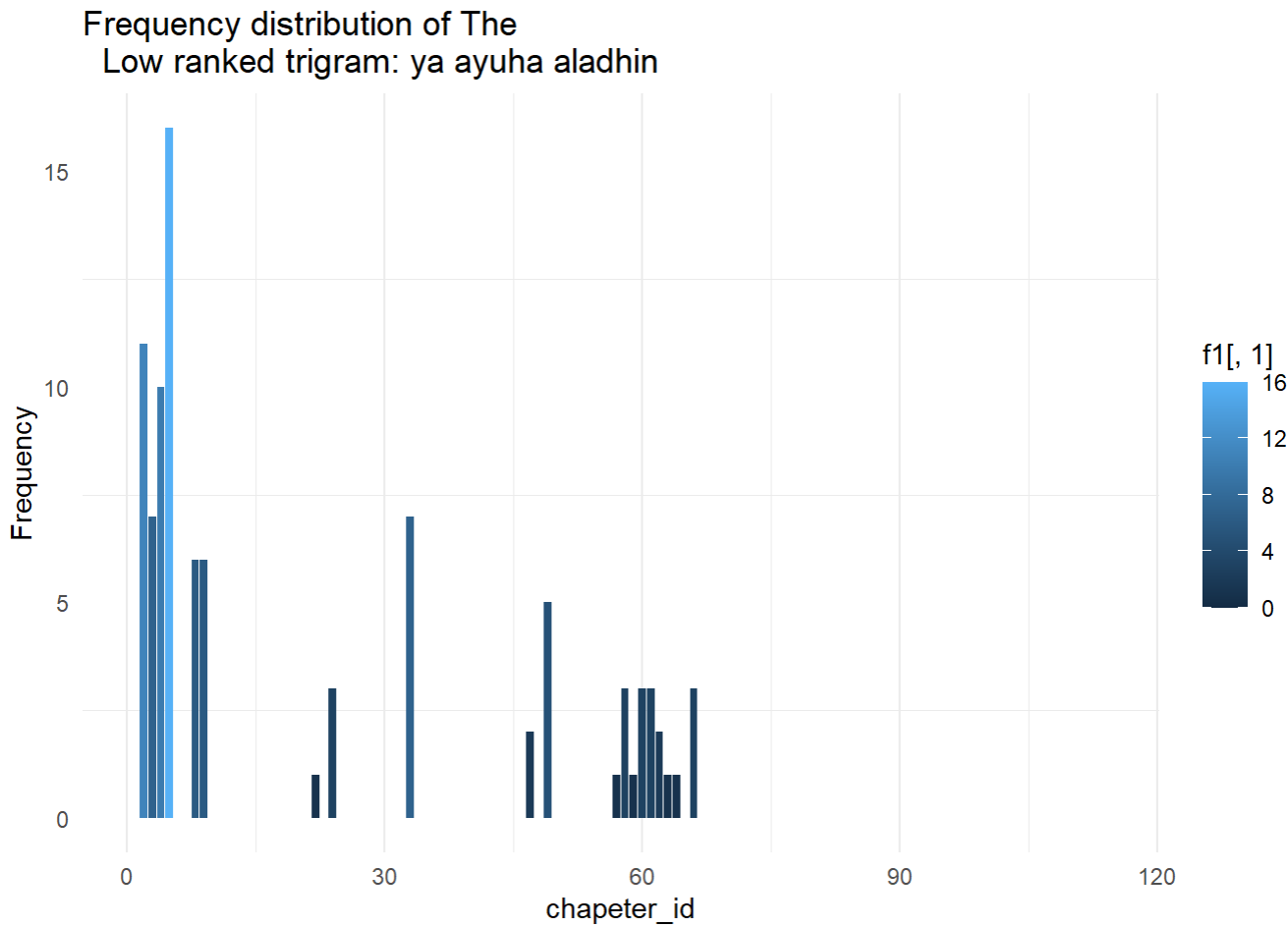
f=as.data.frame(t(f))
f=cbind(f,chapter_id=1:114)

fl=as.data.frame(t(fl))
fl=cbind(fl,chapter_id=1:114)

#Frequency distribution of The top ranked word
ggplot(f,aes(fill=f[,1],x=chapter_id,y=f[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapeter_id",title = "Frequency distribution of The top ranked trigram: ala rabikuma tukadhiban") +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
ggplot(f1,aes(fill=f1[,1],x=chapter_id,y=f1[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapter_id",title = "Frequency distribution of The
  Low ranked trigram: ya ayuha aladhin")+
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

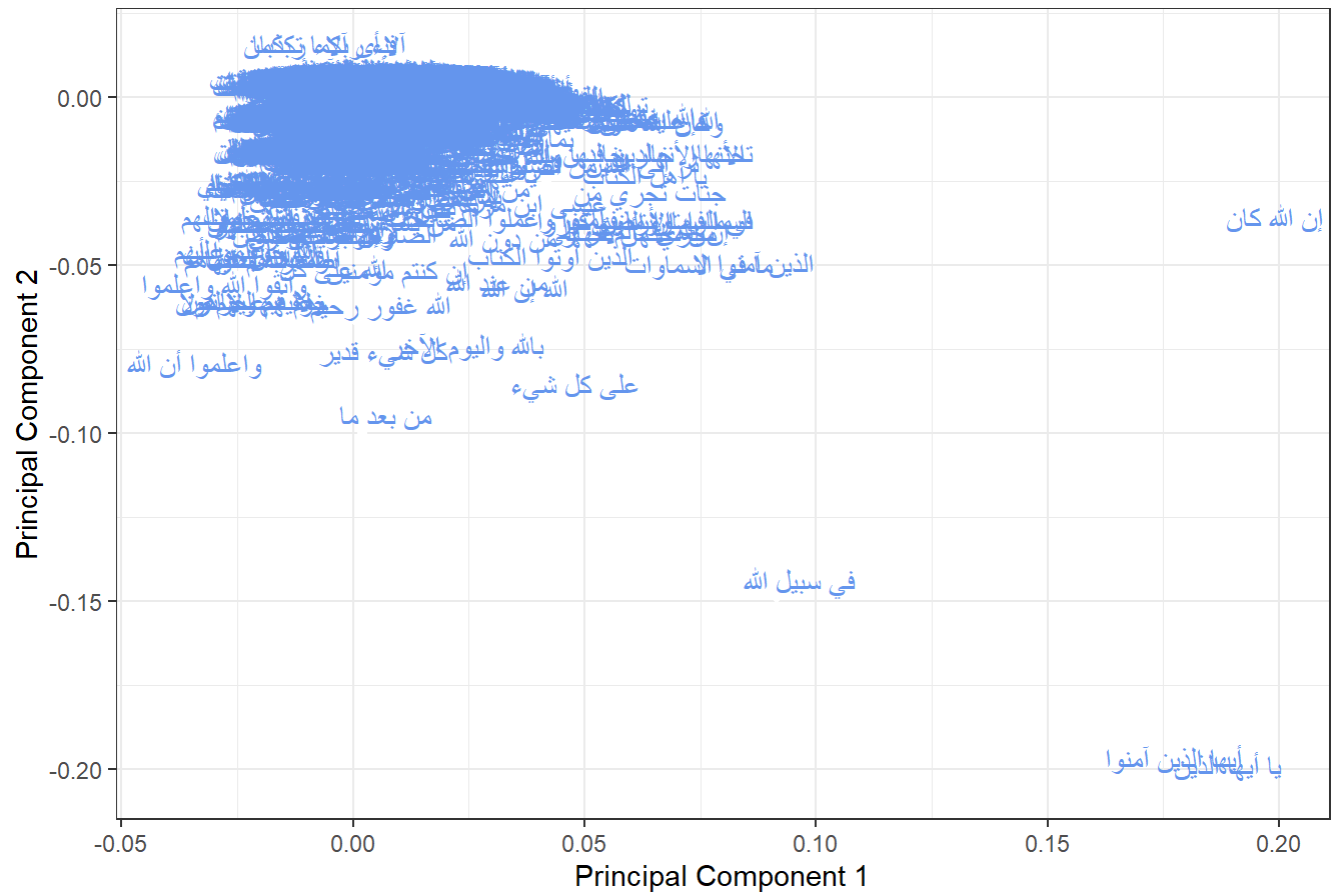


*#To Detect multidimensional outliers, if any, from the projection of
#the data:*

```
pc12=data.frame(pca1$rotation[,c(1,2)])

ggplot(pc12,
  aes(x=PC2,
    y=PC1)) +
  geom_point(color="white") +
  theme_bw() +
  ggtitle('First 2 prinipal components') +
  xlab("Principal Component 1") +
  ylab("Principal Component 2")+
  geom_text(label=rownames(pc12),
    nudge_x = 0.005,
    nudge_y = 0.005,
    size=4,colour="cornflowerblue")
```


First 2 principal components



#To ranke of the Holy Qur'an trigrams in englis:

```
book_id=rep(1,nrow(quran_en_sahih))
quran_en_sahih1=cbind(book_id,quran_en_sahih)
head(quran_en_sahih1)
```

##	book_id	surah_id	ayah_id	surah_title_ar	surah_title_en	surah_title_en_trans
## 1	1	1	1	الفاتحة	Al-Faatiha	The Opening
## 2	1	1	2	الفاتحة	Al-Faatiha	The Opening
## 3	1	1	3	الفاتحة	Al-Faatiha	The Opening
## 4	1	1	4	الفاتحة	Al-Faatiha	The Opening
## 5	1	1	5	الفاتحة	Al-Faatiha	The Opening
## 6	1	1	6	الفاتحة	Al-Faatiha	The Opening
##	revelation_type					
## 1	Meccan					
## 2	Meccan					
## 3	Meccan					
## 4	Meccan					
## 5	Meccan					
## 6	Meccan					
##	text surah					
## 1	In the name of Allah, the Entirely Merciful, the Especially Merciful.					1
## 2	[All] praise is [due] to Allah, Lord of the worlds -					1
## 3	The Entirely Merciful, the Especially Merciful,					1

```
## 4 Sovereign of the Day of Recompense. 1
## 5 It is You we worship and You we ask for help. 1
## 6 Guide us to the straight path - 1
## ayah ayah_title juz manzil page hizb_quarter sajda sajda_id sajda_recommended
## 1 1 1:1 1 1 1 1 FALSE NA NA
## 2 2 1:2 1 1 1 1 FALSE NA NA
## 3 3 1:3 1 1 1 1 FALSE NA NA
## 4 4 1:4 1 1 1 1 FALSE NA NA
## 5 5 1:5 1 1 1 1 FALSE NA NA
## 6 6 1:6 1 1 1 1 FALSE NA NA
## sajda_obligatory
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
```

```
quran_words <- quran_en_sahih1 %>%
  unnest_tokens(trigram, text, token = "ngrams", n = 3) %>%
  count(book_id, trigram, sort = TRUE)

colnames(quran_words)=c("book_id","trigram","freq_in_book")

head(quran_words)
```

```
## book_id trigram freq_in_book
## 1 1 the heavens and 170
## 2 1 and do not 169
## 3 1 indeed allah is 157
## 4 1 and allah is 149
## 5 1 who have believed 147
## 6 1 and those who 137
```

```
chapter_words <- quran_en_sahih1 %>%
  unnest_tokens(trigram, text, token = "ngrams", n = 3) %>%
  count(surah_title_en, trigram, sort = TRUE)

colnames(chapter_words)=c("surah_title_en","trigram","freq_in_chapter")

chapter_words=na.omit(chapter_words)
head(chapter_words)
```

```
## surah_title_en trigram freq_in_chapter
## 1 Ar-Rahmaan of your lord 33
## 2 Ar-Rahmaan favors of your 31
## 3 Ar-Rahmaan lord would you 31
## 4 Ar-Rahmaan of the favors 31
## 5 Ar-Rahmaan so which of 31
## 6 Ar-Rahmaan the favors of 31
```

```
quran_words=left_join(quran_words,chapter_words)
```

```
## Joining, by = "trigram"
```

```
#arabic_stopwords = data.frame(word = removeStopWords("ÓàÇã")$arabicStopwordList)

#book_words=anti_join(book_words,arabic_stopwords, by = "word")
head(quran_words)
```

##	book_id	trigram	freq_in_book	surah_title_en	freq_in_chapter
## 1	1	the heavens and	170	Aal-i-Imraan	9
## 2	1	the heavens and	170	Al-Baqara	8
## 3	1	the heavens and	170	Al-An'aam	8
## 4	1	the heavens and	170	An-Nisaa	6
## 5	1	the heavens and	170	Yunus	6
## 6	1	the heavens and	170	Az-Zumar	6

```
token_list=as.data.frame(table(chapter_words$surah_title_en,chapter_words$trigram))
token_list=token_list[,-3]
colnames(token_list)=c("surah_title_en","trigram")
head(token_list)
```

##	surah_title_en	trigram
## 1	Al-Faatiha	a bad condition
## 2	Al-Baqara	a bad condition
## 3	Aal-i-Imraan	a bad condition
## 4	An-Nisaa	a bad condition
## 5	Al-Maaaida	a bad condition
## 6	Al-An'aam	a bad condition

```
#book_words=mutate(book_words,ratio=book_words$freq_in_chapter/
# book_words$freq_in_book)
head(quran_words)
```

##	book_id	trigram	freq_in_book	surah_title_en	freq_in_chapter
## 1	1	the heavens and	170	Aal-i-Imraan	9
## 2	1	the heavens and	170	Al-Baqara	8
## 3	1	the heavens and	170	Al-An'aam	8
## 4	1	the heavens and	170	An-Nisaa	6
## 5	1	the heavens and	170	Yunus	6
## 6	1	the heavens and	170	Az-Zumar	6

```
quran_words=left_join(token_list,quran_words,by=c("surah_title_en","trigram"))

quran_words[is.na(quran_words)]=0
head(quran_words)
```

##	surah	title	en	trigram	book	id	freq	in	book	freq	in	chapter
----	-------	-------	----	---------	------	----	------	----	------	------	----	---------

##	1	Al-Faatiha	a bad condition	0	0	0
##	2	Al-Baqara	a bad condition	0	0	0
##	3	Aal-i-Imraan	a bad condition	0	0	0
##	4	An-Nisaa	a bad condition	0	0	0
##	5	Al-Maaida	a bad condition	0	0	0
##	6	Al-An'aam	a bad condition	0	0	0

```
r=unique(quran_words$trigram)
c=unique(quran_en_sahih$surah_title_en)

dtrigram=matrix(quran_words$freq_in_chapter,nrow =nrow(quran_words)/114 ,
ncol = 114,byrow = T,dimnames = list(r,c))
```

```
head(dtrigram)
```

##	Al-Faatiha	Al-Baqara	Aal-i-Imraan	An-Nisaa	Al-Maaida	Al-An'aam		
## a bad condition	0	0	0	0	0	0		
## a bad omen	0	0	0	0	0	0		
## a bad tree	0	0	0	0	0	0		
## a bad wind	0	0	0	0	0	0		
## a bad word	0	0	0	0	0	0		
## a band of	0	0	0	0	0	0		
##	Al-A'raaf	Al-Anfaal	At-Tawba	Yunus	Hud	Yusuf	Ar-Ra'd	Ibrahim
## a bad condition	1	0	0	0	0	0	0	0
## a bad omen	0	0	0	0	0	0	0	0
## a bad tree	0	0	0	0	0	0	0	1
## a bad wind	0	0	0	0	0	0	0	0
## a bad word	0	0	0	0	0	0	0	1
## a band of	0	0	0	0	0	0	0	0
##	Al-Hijr	An-Nahl	Al-Israa	Al-Kahf	Maryam	Taa-Haa	Al-Anbiyaa	
## a bad condition	0	0	0	0	0	0	0	
## a bad omen	0	0	0	0	0	0	0	
## a bad tree	0	0	0	0	0	0	0	
## a bad wind	0	0	0	0	0	0	0	
## a bad word	0	0	0	0	0	0	0	
## a band of	0	0	0	0	0	0	0	
##	Al-Hajj	Al-Muminoon	An-Noor	Al-Furqaan	Ash-Shu'araa	An-Naml		
## a bad condition	0	0	0	0	0	0		
## a bad omen	0	0	0	0	0	1		
## a bad tree	0	0	0	0	0	0		
## a bad wind	0	0	0	0	0	0		
## a bad word	0	0	0	0	0	0		
## a band of	0	0	0	0	0	0		
##	Al-Qasas	Al-Ankaboot	Ar-Room	Luqman	As-Sajda	Al-Ahzaab	Saba	
## a bad condition	0	0	0	0	0	0	0	
## a bad omen	0	0	0	0	0	0	0	
## a bad tree	0	0	0	0	0	0	0	
## a bad wind	0	0	1	0	0	0	0	
## a bad word	0	0	0	0	0	0	0	
## a band of	1	0	0	0	0	0	0	
##	Faatir	Yaseen	As-Saaffaat	Saad	Az-Zumar	Ghafir	Fussilat	

##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	1	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##	Ash-Shura	Az-Zukhruf	Ad-Dukhaan	Al-Jaathiya	Al-Ahqaf	Muhammad	
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##	Al-Fath	Al-Hujuraat	Qaaf	Adh-Dhaariyat	At-Tur	An-Najm	Al-Qamar
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##	Ar-Rahmaan	Al-Waaqia	Al-Hadid	Al-Mujaadila	Al-Hashr		
##	a bad condition	0	0	0	0	0	
##	a bad omen	0	0	0	0	0	
##	a bad tree	0	0	0	0	0	
##	a bad wind	0	0	0	0	0	
##	a bad word	0	0	0	0	0	
##	a band of	0	0	0	0	0	
##	Al-Mumtahana	As-Saff	Al-Jumu'a	Al-Munaafiqoon	At-Taghaabun		
##	a bad condition	0	0	0	0	0	
##	a bad omen	0	0	0	0	0	
##	a bad tree	0	0	0	0	0	
##	a bad wind	0	0	0	0	0	
##	a bad word	0	0	0	0	0	
##	a band of	0	0	0	0	0	
##	At-Talaag	At-Tahrim	Al-Mulk	Al-Qalam	Al-Haaqqa	Al-Ma'aarij	Nooh
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##	Al-Jinn	Al-Muzzammil	Al-Muddaththir	Al-Qiyaama	Al-Insaan		
##	a bad condition	0	0	0	0	0	
##	a bad omen	0	0	0	0	0	
##	a bad tree	0	0	0	0	0	
##	a bad wind	0	0	0	0	0	
##	a bad word	0	0	0	0	0	
##	a band of	0	0	0	0	0	
##	Al-Mursalaat	An-Naba	An-Naazi'aat	Abasa	At-Takwir	Al-Infitaar	
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0

##	a band of	0	0	0	0	0	0
##		Al-Mutaffifin	Al-Inshiqaaq	Al-Burooj	At-Taariq	Al-A'laa	
##	a bad condition	0	0	0	0	0	
##	a bad omen	0	0	0	0	0	
##	a bad tree	0	0	0	0	0	
##	a bad wind	0	0	0	0	0	
##	a bad word	0	0	0	0	0	
##	a band of	0	0	0	0	0	
##		Al-Ghaashiya	Al-Fajr	Al-Balad	Ash-Shams	Al-Lail	Ad-Dhuhaa
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##		Ash-Sharh	At-Tin	Al-Alaq	Al-Qadr	Al-Bayyina	Az-Zalzala
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##		Al-Aadiyaat	Al-Qaari'a	At-Takaathur	Al-Asr	Al-Humaza	Al-Fil
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##		Quraish	Al-Maa'un	Al-Kawthar	Al-Kaafiroon	An-Nasr	Al-Masad
##	a bad condition	0	0	0	0	0	0
##	a bad omen	0	0	0	0	0	0
##	a bad tree	0	0	0	0	0	0
##	a bad wind	0	0	0	0	0	0
##	a bad word	0	0	0	0	0	0
##	a band of	0	0	0	0	0	0
##		Al-Ikhlaas	Al-Falaq	An-Naas			
##	a bad condition	0	0	0			
##	a bad omen	0	0	0			
##	a bad tree	0	0	0			
##	a bad wind	0	0	0			
##	a bad word	0	0	0			
##	a band of	0	0	0			

```
pca1=prcomp(t(dtrigram))
```

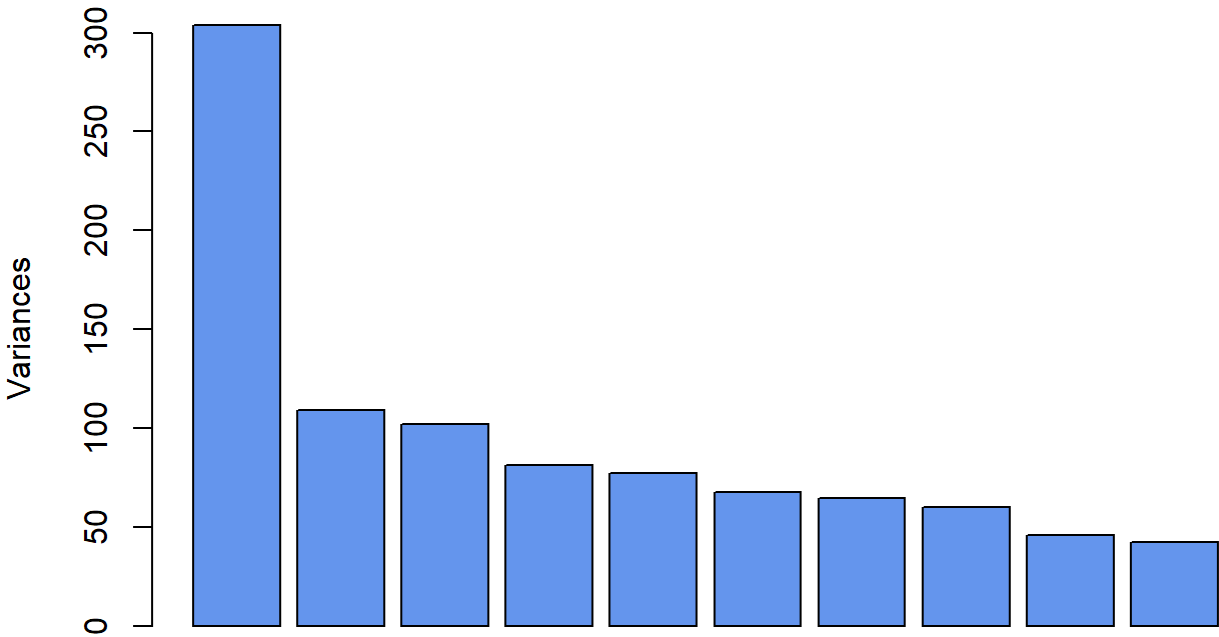
```
pc12=data.frame(pca1$rotation[,c(1,2)])
head(pc12)
```

##		PC1	PC2
##	a bad condition	-7.549614e-04	2.897616e-03
##	a bad omen	1.162882e-04	6.771611e-04

```
## a bad tree      8.659805e-06 1.856774e-05
## a bad wind     -1.569391e-05 2.526025e-04
## a bad word      8.659805e-06 1.856774e-05
## a band of      -8.664279e-05 5.370130e-04
```

```
screepLOT(pca1,col="cornflowerblue",main = "Scree plot of pca of
trigrams distn in englis.")
```

Scree plot of pca of trigrams distn in englis.



```
#To rank the we sort with respect to first principal components:
ranking=pca1[[2]][order(pca1[[2]][,1],decreasing = TRUE),]
head(ranking[,1],10)
```

```
## favors of your      of the favors      lord would you      would you deny      which of the
##      0.008695666      0.008695666      0.008528103      0.008528103      0.007828422
##      so which of your lord would      the favors of      and by the      of your lord
##      0.007708554      0.007493050      0.006126779      0.004336092      0.004133313
```

```
tail(ranking[,1],10)
```

```
## the heavens and      and those who      you who have      o you who
##      -0.08134103      -0.08583960      -0.10174239      -0.10289080
## who have believed      there is no      of allah and      and do not
##      -0.12432663      -0.12664418      -0.14112357      -0.16958501
```

```
## indeed allah is and allah is
## -0.17049264 -0.21198366
```

```
tail1=tail(ranking,20)
ranked=rownames(ranking)
```

#After ranking we got in top ranked trigrams in English:
head(ranked,100)

## [1]	"favors of your"	"of the favors"	"lord would you"
## [4]	"would you deny"	"which of the"	"so which of"
## [7]	"your lord would"	"the favors of"	"and by the"
## [10]	"of your lord"	"the most merciful"	"can make you"
## [13]	"make you know"	"what can make"	"to the deniers"
## [16]	"woe that day"	"day to the"	"that day to"
## [19]	"and what can"	"as for he"	"day of recompense"
## [22]	"and the mountains"	"that day will"	"indeed the righteous"
## [25]	"the righteous will"	"i swear by"	"is the day"
## [28]	"name of your"	"have you seen"	"by the night"
## [31]	"righteous will be"	"is there any"	"faces that day"
## [34]	"enter to burn"	"from the evil"	"any who will"
## [37]	"my punishment and"	"punishment and warning"	"there any who"
## [40]	"who will remember"	"swear by the"	"knowledge of the"
## [43]	"of the right"	"so they are"	"night when it"
## [46]	"the night when"	"day will be"	"in both of"
## [49]	"the mountains are"	"and the heaven"	"word of a"
## [52]	"you seen the"	"for he who"	"to his lord"
## [55]	"indeed the qur'an"	"the mountains will"	"made the qur'an"
## [58]	"they are going"	"of the left"	"the striking calamity"
## [61]	"some faces that"	"so as for"	"servants of allah"
## [64]	"the heaven is"	"by the sky"	"upon us is"
## [67]	"will enter to"	"will be told"	"a cup of"
## [70]	"cup of wine"	"and how severe"	"certainly made the"
## [73]	"easy for remembrance"	"for remembrance so"	"how severe were"
## [76]	"qur'an easy for"	"remembrance so is"	"severe were my"
## [79]	"the qur'an easy"	"were my punishment"	"will be like"
## [82]	"reached you the"	"who is given"	"so i swear"
## [85]	"has there reached"	"there reached you"	"when the sky"
## [88]	"used to deny"	"their eyes humbled"	"among us are"
## [91]	"and among us"	"exalt the name"	"lord the most"
## [94]	"indeed we thus"	"thus reward the"	"we thus reward"
## [97]	"that day is"	"is we who"	"it is we"
## [100]	"they are promised"		

#After ranking we got in last ranked trigrams in English:
tail(ranked,100)

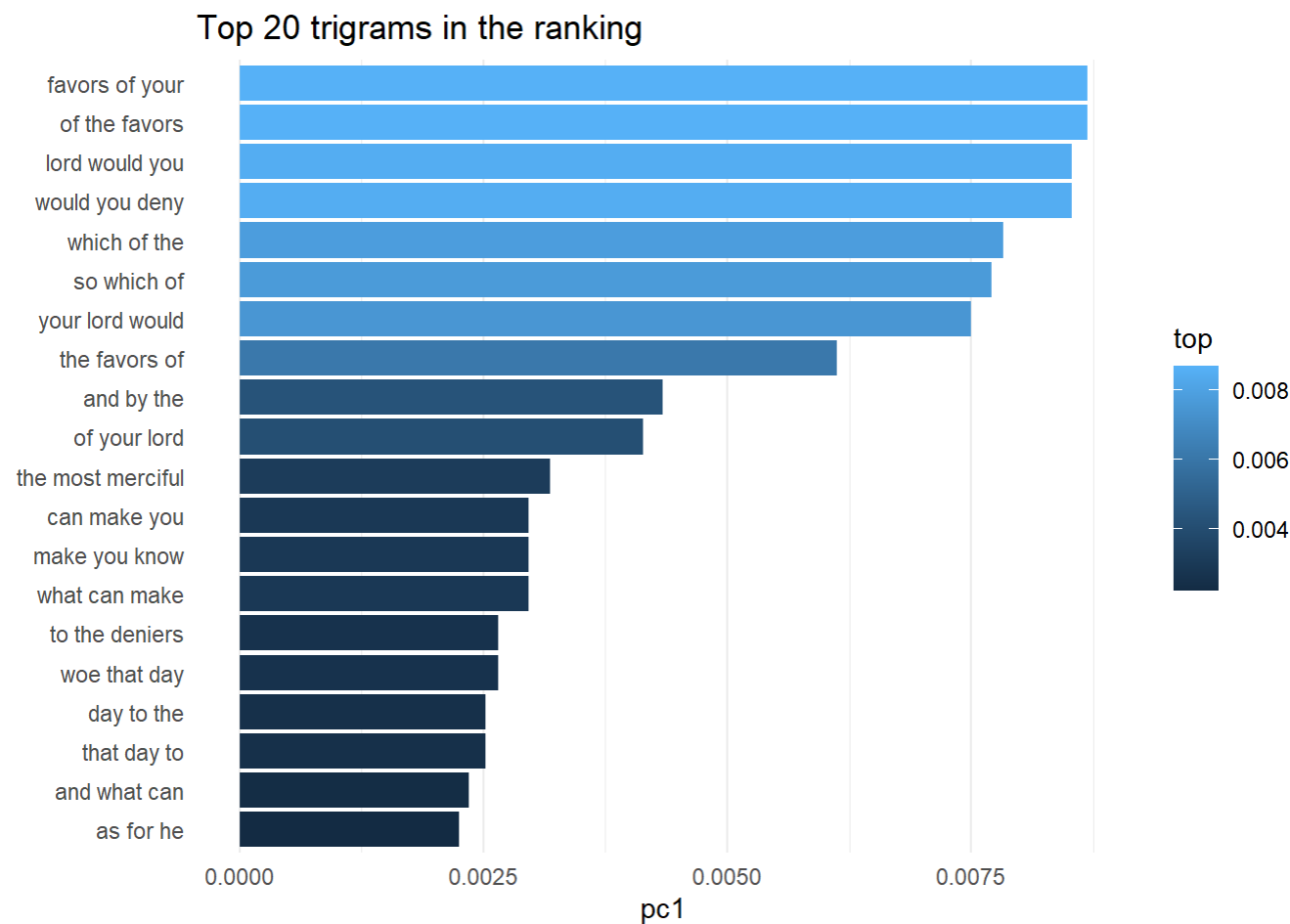
## [1]	"you should be"	"in this world"	"from allah and"
## [4]	"a people who"	"allah and allah"	"and the last"
## [7]	"it is not"	"then indeed allah"	"the people and"
## [10]	"that which is"	"of what they"	"is he who"

##	[13]	"of the people"	"revealed to you"	"them nor will"
##	[16]	"is in the"	"in the cause"	"you do not"
##	[19]	"blame upon you"	"they will abide"	"way of allah"
##	[22]	"heavens and the"	"for you and"	"the last day"
##	[25]	"those who were"	"fear allah and"	"said to them"
##	[28]	"a painful punishment"	"who are the"	"over all things"
##	[31]	"that you may"	"exalted in might"	"of the fire"
##	[34]	"allah is ever"	"they will not"	"nor will they"
##	[37]	"day of resurrection"	"and know that"	"if you are"
##	[40]	"but if you"	"and the earth"	"people of the"
##	[43]	"the day of"	"the way of"	"those who have"
##	[46]	"the earth and"	"is knowing of"	"children of israel"
##	[49]	"you will be"	"but if they"	"is no blame"
##	[52]	"no blame upon"	"the people of"	"of those who"
##	[55]	"and you will"	"on the day"	"he wills and"
##	[58]	"if you should"	"will not be"	"cause of allah"
##	[61]	"what allah has"	"and fear allah"	"those are the"
##	[64]	"those who believe"	"and they will"	"allah and his"
##	[67]	"allah is forgiving"	"the cause of"	"from your lord"
##	[70]	"allah is knowing"	"allah and the"	"that allah is"
##	[73]	"and recall when"	"of what you"	"know that allah"
##	[76]	"for those who"	"is forgiving and"	"of the scripture"
##	[79]	"forgiving and merciful"	"what you do"	"come to you"
##	[82]	"allah does not"	"and if you"	"indeed those who"
##	[85]	"those who are"	"in allah and"	"those who disbelieve"
##	[88]	"you o muhammad"	"and it is"	"whom he wills"
##	[91]	"the heavens and"	"and those who"	"you who have"
##	[94]	"o you who"	"who have believed"	"there is no"
##	[97]	"of allah and"	"and do not"	"indeed allah is"
##	[100]	"and allah is"		

#To plot Top 20 words in the ranking

```
ranking1=data.frame(x1=rownames(ranking)[1:20], top=ranking[1:20,1])

ggplot(ranking1, aes(x = fct_rev(reorder(x1,-top)),
                    y = top, fill = top)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Top 20 trigrams in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

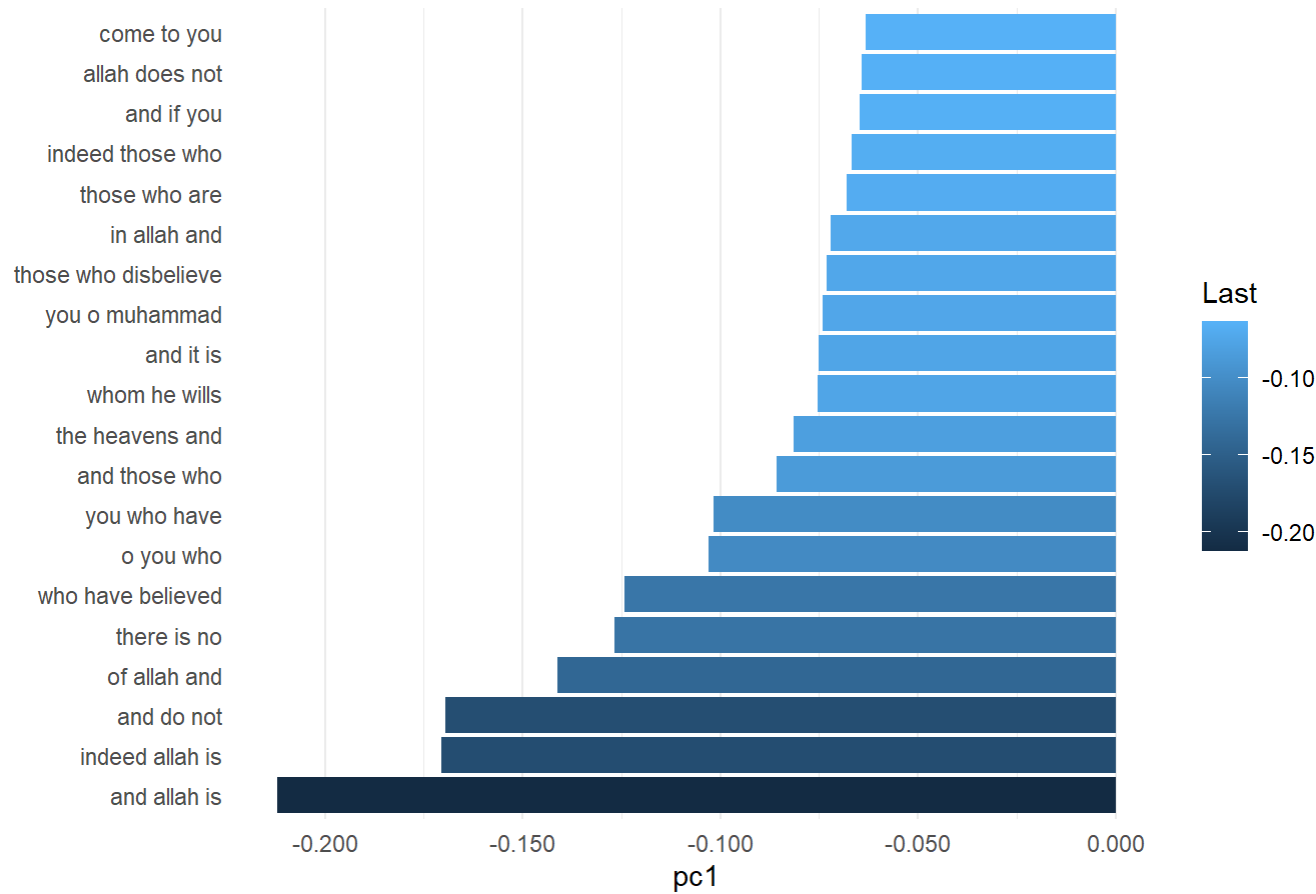


#To plot Last 20 words in the ranking

```
ranking2=data.frame(x1=rownames(tail1),Last=tail1[,1])

ggplot(ranking2, aes(x = fct_rev(reorder(x1,-Last)),
                    y = Last, fill = Last)) +
  geom_col() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "pc1", x = NULL,title = "Last 20 trigrams in the ranking") +
  coord_flip() +
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

Last 20 trigrams in the ranking



```
dtrigram=as.data.frame(dtrigram)

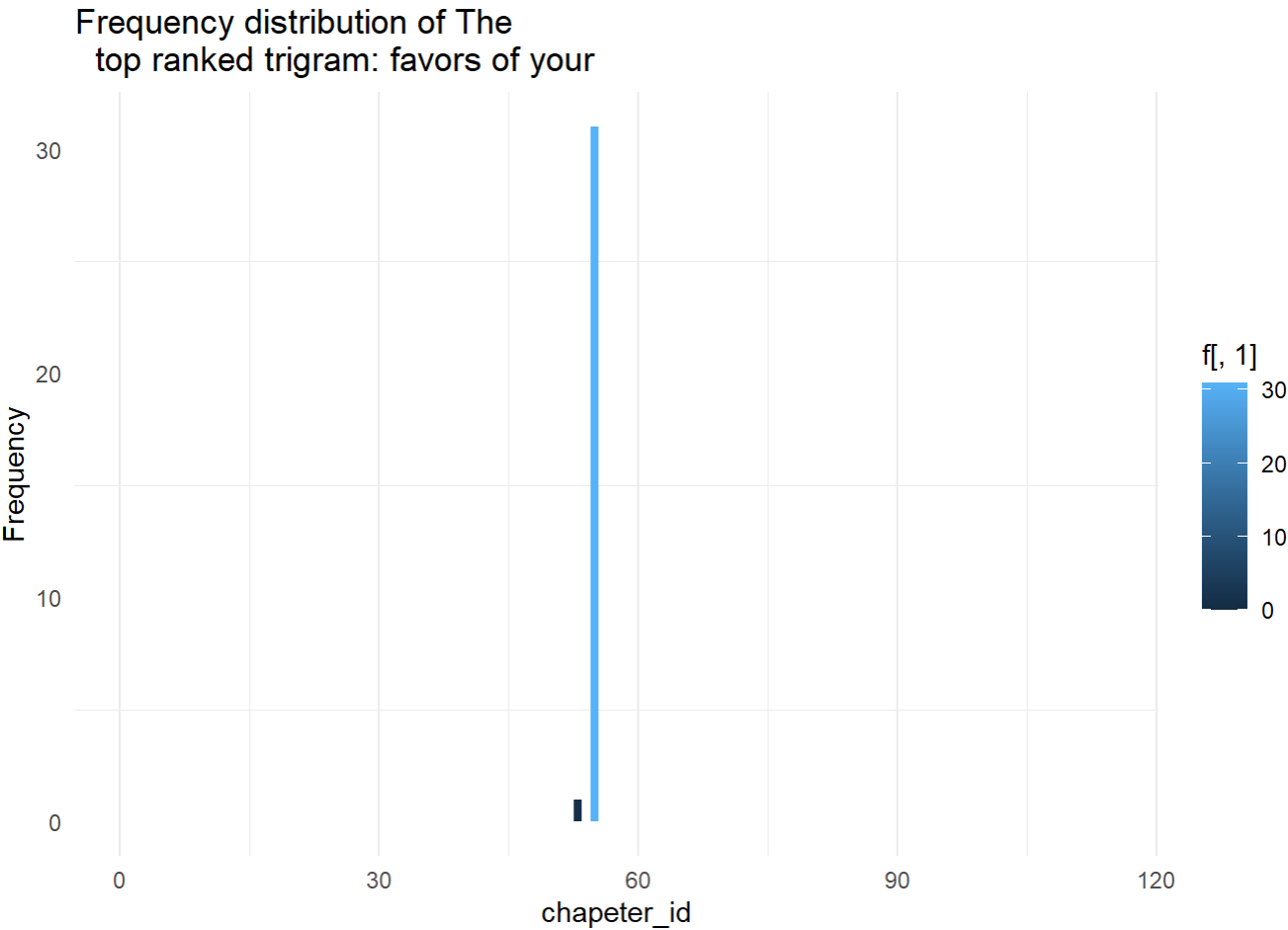
f=filter(dtrigram ,rownames(dtrigram) %in% head(ranked,1))
fl=filter(dtrigram ,rownames(dtrigram) %in% tail(ranked,1))

#view(t(f))
#view(t(fl))

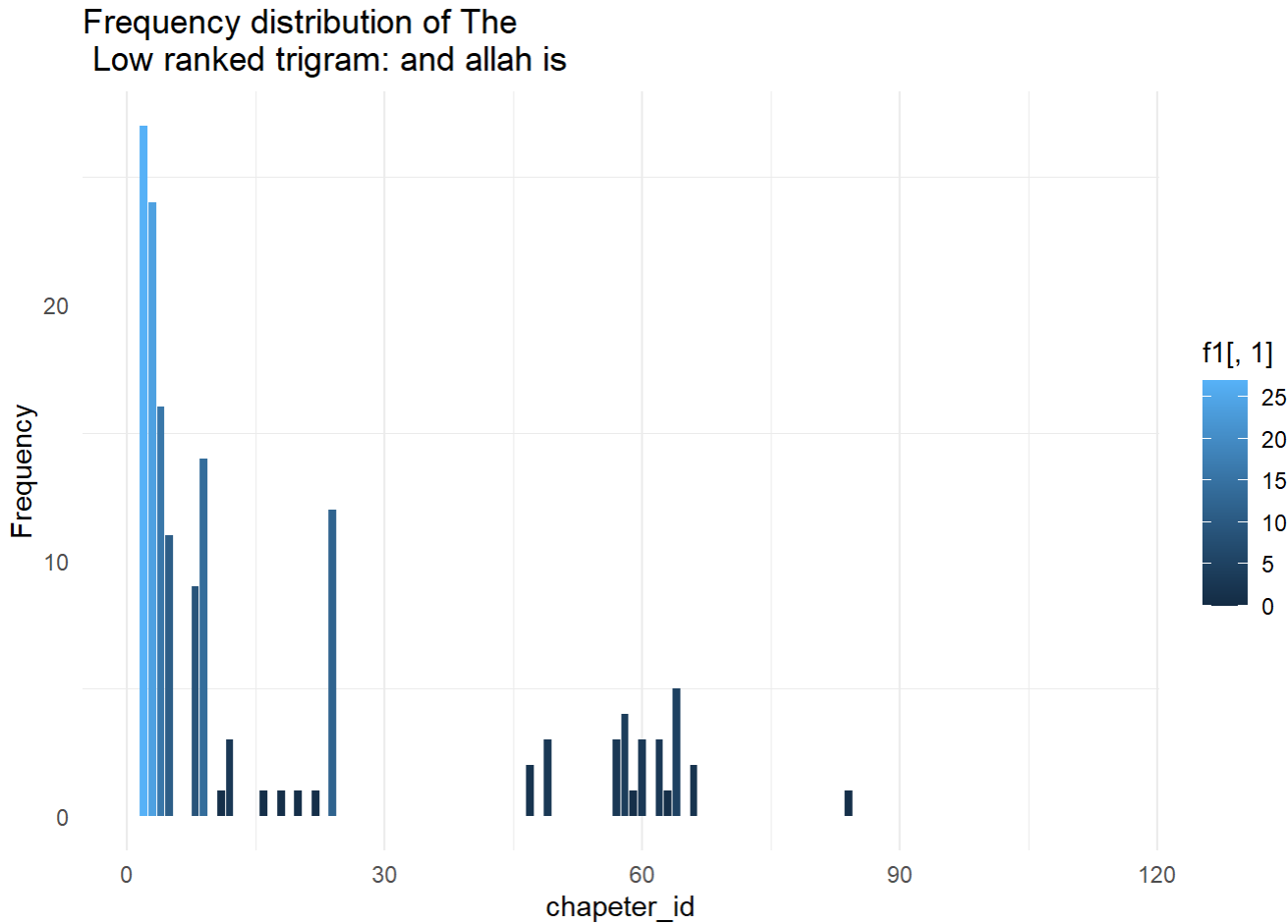
f=as.data.frame(t(f))
f=cbind(f,chapter_id=1:114)

fl=as.data.frame(t(fl))
fl=cbind(fl,chapter_id=1:114)

#Frequency distribution of The top ranked word
ggplot(f,aes(fill=fl[,1],x=chapter_id,y=f[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapter_id",title = "Frequency distribution of The
top ranked trigram: favors of your")+
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```



```
ggplot(f1,aes(fill=f1[,1],x=chapter_id,y=f1[,1]))+
  geom_col()+
  labs(y = "Frequency", x ="chapter_id",title = "Frequency distribution of The
Low ranked trigram: and allah is")+
  theme_minimal() +
  theme(panel.grid.major.y = element_blank())
```

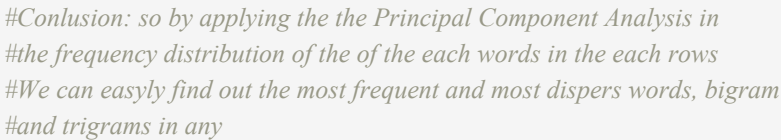


*#To Detect multidimensional outliers, if any, from the projection of
#the data:*

```
pc12=data.frame(pca1$rotation[,c(1,2)])

ggplot(pc12,
  aes(x=PC1,
    y=PC2)) +
  geom_point(color="white") +
  theme_bw() +
  ggtitle('First 2 prinipal components') +
  xlab("Principal Component 1") +
  ylab("Principal Component 2")+
  geom_text(label=rownames(pc12),
    nudge_x = 0.005,
    nudge_y = 0.005,
    size=4,colour="cornflowerblue")
```

Principal Component 2



"الله" "الذين" "والله" "أمنوا" "فإن" "كان" [1]
"قال" "قل" "الكتاب" "الناس" "قالوا" "بالله" [7]
"كنتم" "الأرض" "أيها" "كفروا" "كل" "فلا" [13]
"شيء" "كانوا" "أنزل" "أولئك" "والذين" "علیم" [19]
"خير" "سبیل" "الدنیا" "وهو" "لله" "يشاء" [25]
"السموات" "عذاب" "النار" "الحق" "ربنا" "أنفسهم" [31]
"إبراهيم" "وأنتم" "منه" "موسی" "وكان" "ربكم" [37]
"فقد" "يوم" "لعلكم" "المؤمنین" "تعملون" "منها" [43]
"واقتوا" "ربهم" "یؤمنون" "الصلاة" "أهل" "والأرض" [49]

[1] "الحاقة" يسرا "لقلول" يره "فضة" القدر " [1]
[7] "إنس" الميزان "ينبغي" حسابا "القارة" كيدا " [7]
[13] "الجهال" رجم "سلاما" ممنون "وبنين" صلصال " [13]
[19] "يتساءلون" أفرايم "نبيا" ولبنا "الأراك" المرسلون " [19]

```
## [25] "المأوى" "آخر" "أدبر" "نادى" "عذابي" "سقر"
## [31] "يغني" "أواب" "ترايا" "ورب" "الصراط" "يخشى"
## [37] "يكدب" "ونذر" "مصبحين" "فذكر" "الشمال" "بالنذر"
## [43] "جند" "بسم" "أينا" "المسكين" "يسرنا" "سرر"
## [49] "باسم" "وكتا" "فكانت" "إإذا" "راضية" "أيحسب"
```

#After ranking we got in top ranked for Englis:

```
## [1] "allah" "people" "lord" "fear" "believed"
## [6] "punishment" "messenger" "knowing" "believers" "earth"
## [11] "day" "muhammad" "scripture" "revealed" "truth"
## [16] "disbelievers" "disbelieve" "fire" "women" "prayer"
## [21] "reward" "merciful" "wills" "righteous" "religion"
## [26] "hearts" "life" "verses" "heavens" "remember"
## [31] "moses" "wise" "abraham" "forgiving" "children"
## [36] "spend" "messengers" "worship" "guided" "fight"
## [41] "mercy" "follow" "guidance" "evil" "angels"
## [46] "death" "sin" "disbelieved" "world" "signs"
```

#After ranking we got in last ranked for Englis:

```
## [1] "prays" "leg" "scorner" "zaqqum"
## [5] "balance" "substance" "multitudes" "severely"
## [9] "expecting" "comer" "planning" "columns"
## [13] "crusher" "humanity" "sparks" "firewood"
## [17] "hot" "disintegrated" "swimming" "observing"
## [21] "repel" "uninterrupted" "foul" "strengthened"
## [25] "contents" "blazing" "pairs" "screaming"
## [29] "pearls" "walks" "ruined" "vessels"
## [33] "enumerated" "scorching" "inscribed" "poet"
## [37] "cool" "ahead" "shout" "stakes"
## [41] "glances" "brocade" "extent" "bitter"
## [45] "thrones" "deprived" "setting" "boys"
## [49] "constructed" "mad" "happiness" "assembled"
```