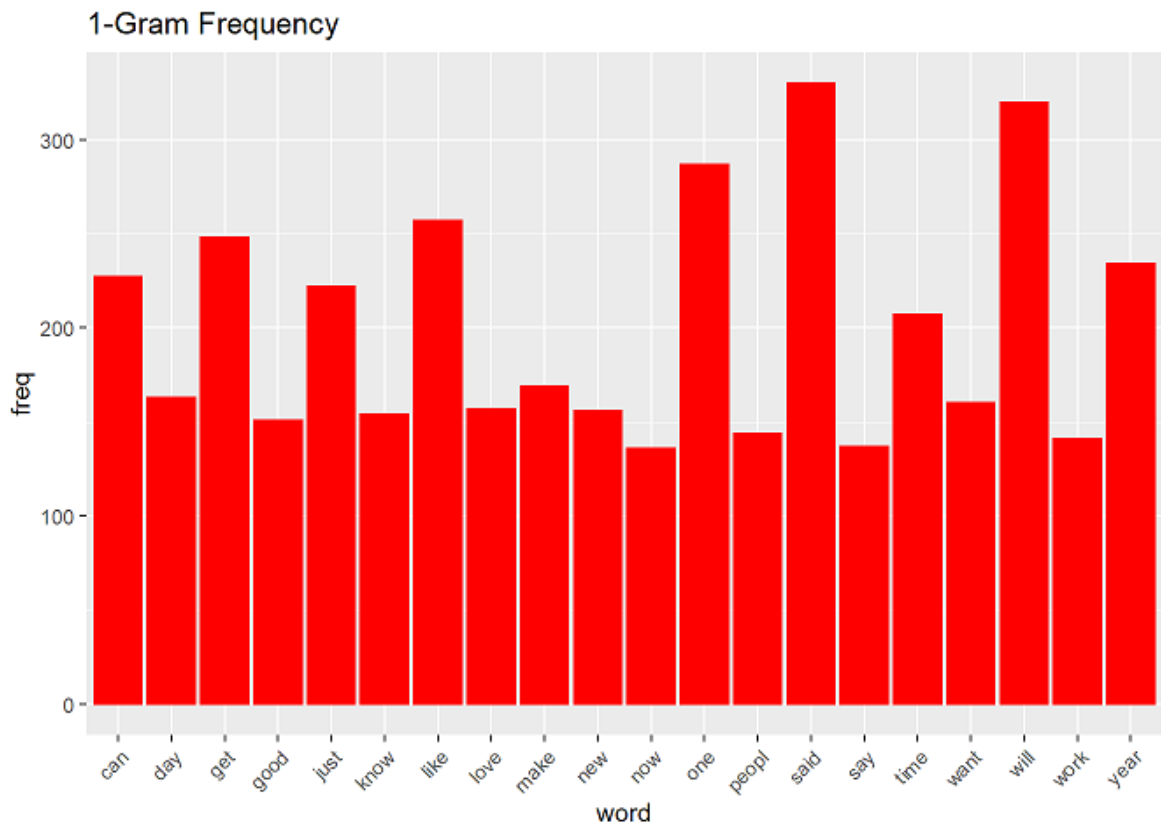


# Data Exploration

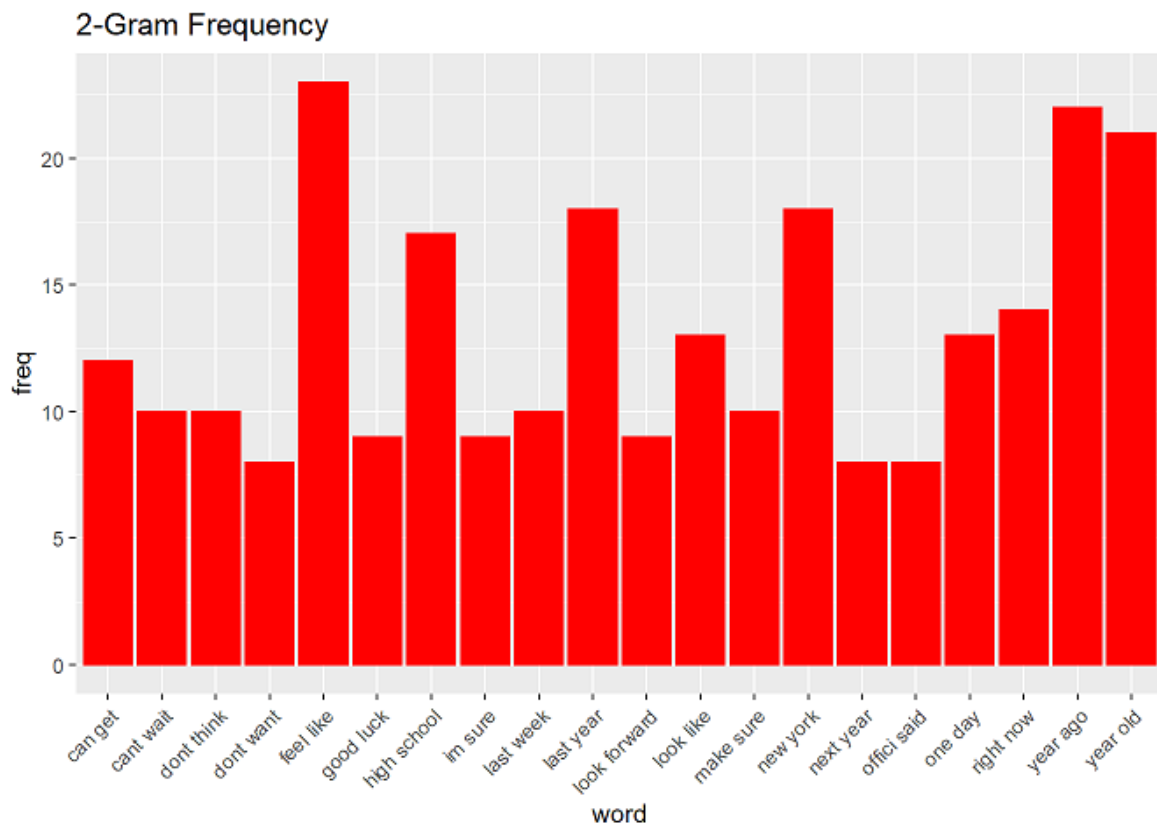
## 1-Gram Frequency

```
freq1 <- rowSums(as.matrix(dtm1))  
freq1 <- sort(freq1, decreasing = TRUE)  
dfFreq1 <- data.frame(word = names(freq1), freq=freq1)  
ggplot(dfFreq1[1:20, ], aes(word, freq)) +  
  geom_bar(stat="identity", fill="red", colour="red") +  
  theme(axis.text.x=element_text(angle=45, hjust=1)) + ggtitle("1-Gram Fr  
equencey")
```



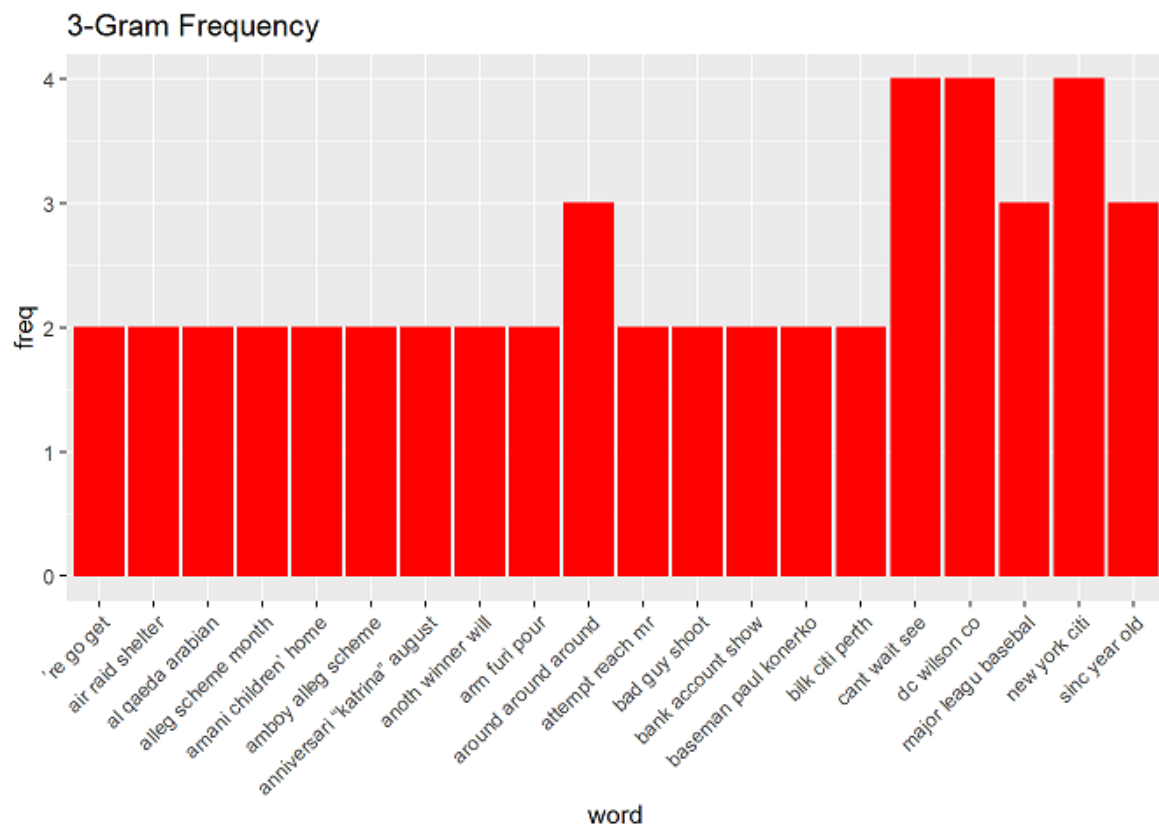
## 2-Gram Frequency

```
freq2 <- rowSums(as.matrix(dtm2))  
freq2 <- sort(freq2, decreasing = TRUE)  
dfFreq2 <- data.frame(word = names(freq2), freq=freq2)  
ggplot(dfFreq2[1:20, ], aes(word, freq)) +  
  geom_bar(stat="identity", fill="red", colour="red") +  
  theme(axis.text.x=element_text(angle=45, hjust=1)) + ggtitle("2-Gram Frequency")
```



### 3-Gram Frequency

```
freq3 <- rowSums(as.matrix(dtm3))  
freq3 <- sort(freq3, decreasing = TRUE)  
dfFreq3 <- data.frame(word = names(freq3), freq=freq3)  
ggplot(dfFreq3[1:20, ], aes(word, freq)) +  
  geom_bar(stat="identity", fill="red", colour="red") +  
  theme(axis.text.x=element_text(angle=45, hjust=1)) + ggtitle("3-Gram Frequency")
```



## Future work

The goal is to create a predictive model which predicts the most probable words to follow an input from the user. This model will be evaluated and deployed as a shiny application.