

text mining-4

20143311 김희수

A: Data Preparation

B : Text Wording

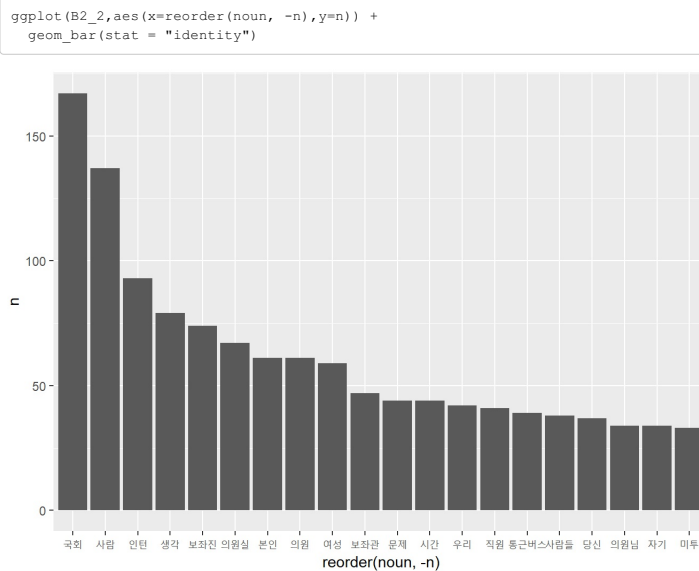
```
B1 <- A1$post %>%
  SimplePos09 %>%
  melt %>%
  as_tibble %>%
  select(3, 1)

B1_An <- A1_An %>%
  SimplePos09 %>%
  melt %>%
  as_tibble %>%
  select(3, 1)

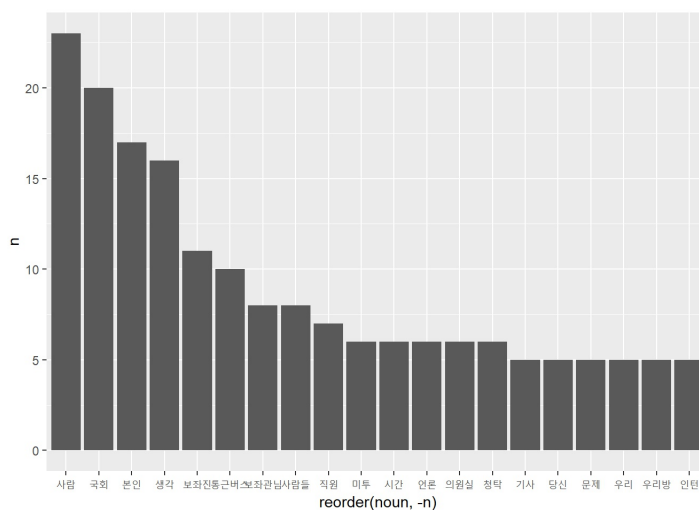
# `명사에 관하여`
B2 <- B1 %>%
  mutate(noun=str_match(value, '([가-힣])+ /N')[,2]) %>%
  na.omit %>%
  filter(str_length(noun)>=2) %>%
  count(noun, sort=TRUE)

B2_An <- B1_An %>%
  mutate(noun=str_match(value, '([가-힣])+ /N')[,2]) %>%
  na.omit %>%
  filter(str_length(noun)>=2) %>%
  count(noun, sort=TRUE)

B2_2 <- head(B2,20)
B2_2_An <- head(B2_An,20)
```



```
ggplot(B2_2_An,aes(x=reorder(noun, -n),y=n)) +
  geom_bar(stat = "identity")
```

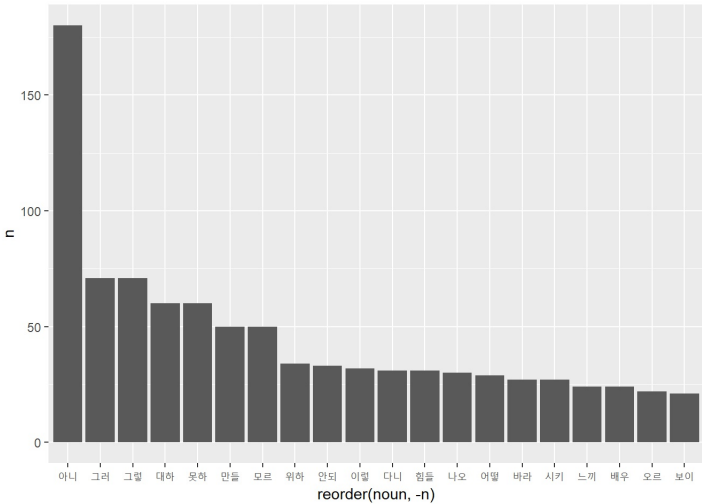


'동사에 관하여'

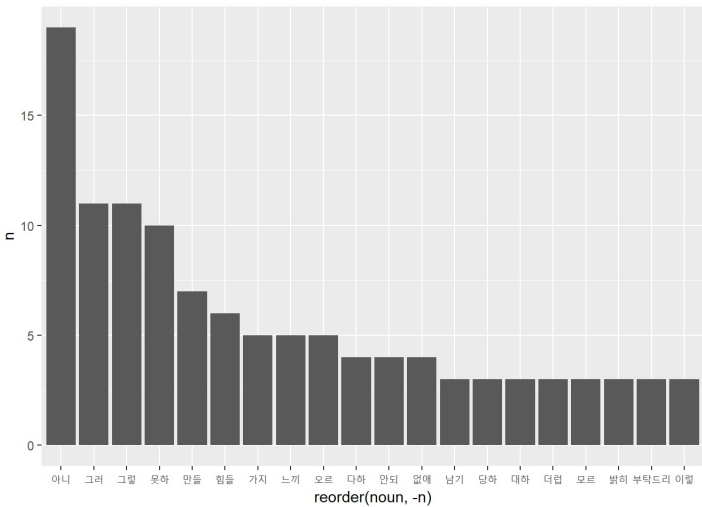
```
B3 <- B1 %>%
  mutate(noun=str_match(value, '([가-힣]+)/P')[,2]) %>%
  na.omit %>%
  filter(str_length(noun)>=2) %>%
  count(noun, sort=TRUE)
B3_An <- B1_An %>%
  mutate(noun=str_match(value, '([가-힣]+)/P')[,2]) %>%
  na.omit %>%
  filter(str_length(noun)>=2) %>%
  count(noun, sort=TRUE)

B3_2 <- head(B3,20)
B3_2_An <- head(B3_An,20)

ggplot(B3_2,aes(x=reorder(noun, -n),
                    y=n)) +
  geom_bar(stat = "identity")
```



```
ggplot(B3_2_An,aes(x=reorder(noun, -n),
                    y=n)) +
  geom_bar(stat = "identity")
```



C : Emotional Analysis

```
{r Emotional Analysis,results='asis'} #단어에 대한 감정 부여 (긍정,부정,중립) C1 <- B2[1:30,] C1
emotion <- " 긍정 " C1[1,3] <- " 부정 " C1[2,3] <- " 긍정 " C1[3,3] <- " 부정 "
emotion <- factor(B1$emotion,levels =c("긍정","부정"), labels =c("긍정","부정"))

C2 <- C1 %>% plyr::ddply("emotion", summarise, emotion_level = sum(n)/sum(B1$n)*100)

kable(C2)
```

D : Word Cloud

```
wordcloud2(B2[1:30,],fontFamily = '나눔고딕',
            minRotation=0, maxRotation=0)
```



```
wordcloud2(B2_An[1:30,],fontFamily = '나눔고딕',  
            minRotation=0, maxRotation=0)
```

E : Co-Occurrence Network Anlysis