1. inaugural_address.csv를 불러와 분석에 적합하게 전처리한 다음 연설문에서 명사를 추출하세요.

R 코드

```
# 1.
speeches <- raw_speeches %>%
            mutate(value=str_replace_all(value, pattern="[^가-힣]", replacement=" "),
                   value=str_squish(value))

speeches_noun <- speeches %>%
                 unnest_tokens(input = value,
                               output = word,
                               token = extractNoun)

frequency <- speeches_noun %>%
             count(president, word) %>%
             filter(str_length(word) > 1) %>%
             print()
```

R 프로그램 결과

	president	word	n
1	노무현	가난	1
2	노무현	가능	1
3	노무현	가일	1
4	노무현	가지	1
5	노무현	각국	1
6	노무현	갈등	1
7	노무현	갈림길	1
8	노무현	감사	4
9	노무현	강구	2
10	노무현	강국	1
11	노무현	강요	1
12	노무현	강조	1

Showing 1 to 12 of 1,657 entries, 3 total columns

2. TF-IDF를 이용해 각 연설문에서 상대적으로 중요한 단어를 10개씩 추출하세요.

R 코드

```r
# 2.
frequecy <- frequecy %>%
            bind_tf_idf(term = word,
                        document = president,
                        n = n) %>%
            arrange(-tf_idf) %>%
            print()

frequecy %>% filter(president == "문재인")
frequecy %>% filter(president == "박근혜")
frequecy %>% filter(president == "이명박")
frequecy %>% filter(president == "노무현")

# 주요 단어 추출
top10_four <- frequecy %>%
              group_by(president) %>%
              slice_max(tf_idf, n = 10, with_ties = F)
```

2. TF-IDF를 이용해 각 연설문에서 상대적으로 중요한 단어를 10개씩 추출하세요.

R 프로그램 결과

president <fctr>	word <chr>	n <int>	tf <dbl>	idf <dbl>	tf_idf <dbl>	president <fctr>	word <chr>	n <int>	tf <dbl>	idf <dbl>	tf_idf <dbl>
노무현	공식	6	0.016348774	1.3862944	0.022664213	이명박	리더십	6	0.015789474	1.3862944	0.021888858
노무현	비전	6	0.016348774	1.3862944	0.022664213	이명박	당원	4	0.010526316	1.3862944	0.014592572
노무현	정계	6	0.016348774	1.3862944	0.022664213	이명박	동지	4	0.010526316	1.3862944	0.014592572
노무현	권력	9	0.024523161	0.6931472	0.016998160	이명박	일류국가	4	0.010526316	1.3862944	0.014592572
노무현	개편	4	0.010899183	1.3862944	0.015109475	이명박	한나라	7	0.018421053	0.6931472	0.012768501
노무현	국회의원	3	0.008174387	1.3862944	0.011332106	이명박	나라	15	0.039473684	0.2876821	0.011355871
노무현	남북대화	3	0.008174387	1.3862944	0.011332106	이명박	도덕	3	0.007894737	1.3862944	0.010944429
노무현	중리	3	0.008174387	1.3862944	0.011332106	이명박	일하	3	0.007894737	1.3862944	0.010944429
노무현	가운	2	0.005449591	1.3862944	0.007554738	이명박	사랑	5	0.013157895	0.6931472	0.009120358
노무현	개혁	4	0.010899183	0.6931472	0.007554738	이명박	인생	5	0.013157895	0.6931472	0.009120358

president <fctr>	word <chr>	n <int>	tf <dbl>	idf <dbl>	tf_idf <dbl>	president <fctr>	word <chr>	n <int>	tf <dbl>	idf <dbl>	tf_idf <dbl>
문재인	복지국가	8	0.006083650	1.3862944	0.008433730	박근혜	박근혜	8	0.009615385	1.3862944	0.013329753
문재인	여성	6	0.004562738	1.3862944	0.006325297	박근혜	정보	5	0.006009615	1.3862944	0.008331096
문재인	공평	5	0.003802281	1.3862944	0.005271081	박근혜	투명	5	0.006009615	1.3862944	0.008331096
문재인	담쟁이	5	0.003802281	1.3862944	0.005271081	박근혜	행복	23	0.027644231	0.2876821	0.007952750
문재인	대통령의	5	0.003802281	1.3862944	0.005271081	박근혜	교육	9	0.010817308	0.6931472	0.007497986
문재인	보통	5	0.003802281	1.3862944	0.005271081	박근혜	국정운영	4	0.004807692	1.3862944	0.006664877
문재인	상생	5	0.003802281	1.3862944	0.005271081	박근혜	정부	17	0.020432692	0.2876821	0.005878119
문재인	우리나라	10	0.007604563	0.6931472	0.005271081	박근혜	개개인	3	0.003605769	1.3862944	0.004998658
문재인	지방	5	0.003802281	1.3862944	0.005271081	박근혜	개인	3	0.003605769	1.3862944	0.004998658
문재인	확대	10	0.007604563	0.6931472	0.005271081	박근혜	공개	3	0.003605769	1.3862944	0.004998658

3. 각 연설문에서 상대적으로 중요한 단어를 나타낸 막대 그래프를 만드세요.

R 코드

```r
# 3.
# 그래프 순서 정하기
top10_four$president <- factor(top10_four$president,
                               levels = c("문재인", "박근혜", "이명박", "노무현"))

# '나눔고딕' 폰트 적용
library(showtext)
font_add_google(name = "Nanum Gothic", family = "nanumgothic")
showtext_auto()

# 막대 그래프 만들기
top10_four %>% ggplot(aes(x = reorder_within(x=word, by=tf_idf,
                                             within=president),
                          y = tf_idf, fill = president)) +
               geom_col(show.legend = F) +
               coord_flip() +
               facet_wrap(~ president, scales = "free", ncol = 2) +
               scale_x_reordered() +
               labs(x = NULL) +
               theme(text = element_text(family = "nanumgothic"))
```

R 프로그램 결과