

1. speech_park.txt를 불러와 분석에 적합하게 전처리한 다음 연설문에서 명사를 추출하세요.

R 코드

```
# 5-1
setwd('C:/Users/jspar/OneDrive/Documents/학교/전공/텍마')
library(tidyr)
library(tidytext)
library(tidyverse)

raw_speech <- readLines("speech_park.txt", encoding = "UTF-8")

speech <- raw_speech %>%
  str_replace_all("[^가-힣]", " ") %>%
  str_squish() %>%
  as_tibble() %>%
  filter(value!='')

noun_token <- speech %>%
  unnest_tokens(input = value,
                output= word,
                token = extractNoun) %>%
  count(word, sort = T) %>%
  filter(str_length(word) > 1)

noun_token
```

R 프로그램 결과

	word	n
1	국민	72
2	행복	23
3	여러분	20
4	정부	17
5	경제	15
6	신뢰	11
7	국가	10
8	우리	10
9	교육	9
10	사람	9

Showing 1 to 10 of 427 entries, 2 total columns

2. 가장 자주 사용된 단어 20개를 추출하세요.

R 코드

```
# 5-2
word_count <- speech %>%
  unnest_tokens(input = value,
                output = word,
                token = 'words') %>%
  count(word, sort = T) %>%
  filter(str_length(word) > 1) %>%
  head(20)

word_count
```

R 프로그램 결과

	word	n
1	국민	29
2	저는	14
3	있습니다	12
4	함께	12
5	꿈을	10
6	것입니다	8
7	새로운	8
8	있는	8
9	국민행복의	7
10	길을	7

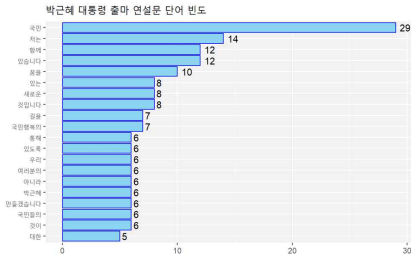
Showing 1 to 10 of 20 entries, 2 total columns

3. 가장 자주 사용된 단어 20개의 빈도를 나타낸 막대 그래프를 만드세요.

R 코드

```
# 5-3
word_count %>%
  ggplot(aes(x=reorder(word, n), y = n)) +
    geom_col(colour='blue', fill='skyblue') +
    coord_flip() +
    geom_text(aes(label = n), hjust = -0.4) +
    labs(title = '박근혜 대통령 출마 연설문 단어 빈도',
         x = NULL, y = NULL)
```

R 프로그램 결과



4. 전처리 하지 않은 연설문에서 연속된 공백을 제거하고 tibble 구조로 변환한 다음 문장 기준으로 토큰화 하세요.

R 코드

```
# 5-4
st_speech <- raw_speech %>%
  str_squish() %>%
  as_tibble() %>%
  unnest_tokens(input = value,
                 output = sentence,
                 token = 'sentences')

st_speech
```

R 프로그램 결과

	sentence
1	존경하는 국민 여러분!
2	저는 오늘, 국민 한 분 한 분의 꿈이 이루어지는 행복한 대한...
3	저 박근혜, 이번 18대 대통령선거 출마를 선언합니다.
4	국민 여러분!
5	저의 삶은 대한민국과 함께 해온 시간이었습니다.
6	우리나라가 가난을 이기고, 꿈을 이뤄가는 위대한 과정을 어...
7	어머니가 출탄에 돌아가신 후, 견딜 수 없는 고통과 어려움 ...
8	그리고 늘 함께 해주시고 힘이 되어 주셨던 국민 여러분이 계...
9	제가 그 막중한 일을 해내고, 고통을 이겨내고 다시 일어설 ...
10	그때부터 제 삶은 완전히 다른 길을 가야했습니다.

Showing 1 to 10 of 119 entries, 1 total columns

5. 연설문에서 "경제"가 사용된 문장을 출력하세요.

R 코드

```
# 5-5
st_speech %>%
  filter(str_detect(string = sentence,
                    pattern = '경제'))
...
```

R 프로그램 결과

	sentence
1	국가는 발전했고, 경제는 성장했다는데, 나의 삶은 나아지지 ...
2	저는 '경제민주화 실현', '일자리 창출', 그리고 '한국형 복지의...
3	국민행복의 길을 열어갈 첫 번째 과제로, 저는 경제민주화를 ...
4	그동안 우리 경제는 효율성을 지나치게 강조하면서 공정성...
5	공정하고 투명한 시장경제 질서를 확립해 경제민주화를 실...
6	정당한 기업활동은 최대한 보장하고 불필요한 규제는 철폐...
7	수출 일변도의 경제구조에서 벗어나 수출과 내수가 동시에 ...
8	복지가 국민 개개인이 가진 자기 역량을 뒷받침하고 끌어내...
9	국민 여러분, 올해는 우리나라 경제발전의 틀을 마련한 경제...
10	저는 국민행복을 위해 '경제민주화-일자리-복지'를 아우르는...
11	50년 전 경제개발 5개년 계획이 산업화의 기적을 이뤄냈듯, '

Showing 1 to 11 of 11 entries, 1 total columns