

D&A Basic Session 과제

4주차

학과 : 빅데이터 경영통계전공

조이름 : 쉽지 않조 바쁘조 왜이렇게 까지 해야하나 싶조

학번 : 20192774

이름 : 신기섭

문제1) 아래 busker란 변수에 저장된 문자열에서 "그대여"가 총 몇 번 나오는지 알아보는 코드를 작성하세요.

코드

```
x<-unlist(strsplit(busker,'\t '));x
a<-grep('그대여',x);a
length(a)
```

결과

```
> x<-unlist(strsplit(busker,'\t '));x
[1] "그대여" "그대여" "그대여" "그대여" "그대여" "오늘은" "우리" "같이" "걸어요" "이" "거리" "밤에" "들려오는" "자장노래" "어떤가요" "오예"
[17] "물랐던" "그대와" "단" "둘이" "손" "장고" "알" "수" "없는" "이" "떨림과" "둘이" "걸어요" "봄바람" "취날리며" "홀날리는"
[33] "벗꽃" "잎이" "울러" "퍼질" "이" "거리" "우우" "둘이" "걸어요" "봄바람" "취날리며" "홀날리는" "벗꽃" "잎이" "울러" "퍼질"
[49] "이" "거리" "우우" "둘이" "걸어요" "오예" "그대여" "우리" "이제" "손" "잡아요" "이" "거리" "마침" "들려오는" "사랑"
[65] "노래" "어떤가요" "오예" "사랑하는" "그대와" "단" "둘이" "손잡고" "알" "수" "없는" "이" "거리" "둘이" "걸어요" "봄바람"
[81] "취날리며" "홀날리는" "벗꽃" "잎이" "울러" "퍼질" "이" "거리" "우우" "둘이" "걸어요" "봄바람" "취날리며" "홀날리는" "벗꽃" "잎이"
[97] "울러" "퍼질" "이" "거리" "우우" "둘이" "걸어요" "바람" "불면" "울림" "기분" "탓에" "나도" "모르게" "바람" "불면"
[113] "저편에서" "그대여" "니" "모습이" "자꾸" "겹쳐" "오" "또" "울림" "기분" "탓에" "나도" "모르게" "바람" "불면" "저편에서"
[129] "그대여" "니" "모습이" "자꾸" "겹쳐" "사랑하는" "연인들이" "망균요" "알" "수" "없는" "친구들이" "망균요" "홀날리는" "벗꽃" "잎이"
[145] "망균요" "좋아요" "봄바람" "취날리며" "홀날리는" "벗꽃" "잎이" "울러" "퍼질" "이" "거리" "우우" "둘이" "걸어요" "봄바람" "취날리며"
[161] "홀날리는" "벗꽃" "잎이" "울러" "퍼질" "이" "거리" "우우" "둘이" "걸어요" "그대여" "그대여" "그대여" "그대여" "그대여"
> a<-grep('그대여',x);a
[1] 1 2 3 4 5 55 114 129 171 172 173 174 175
> length(a)
[1] 13
```

문제2) 아래 text란 변수에는 로미오와 줄리엣의 대본이 저장되어있습니다. 로미오와 줄리엣은 희곡으로 막/장으로 구분되어 있습니다(ACT/Scene). 각 장(Scene)의 장소를 추출하여 scenes라는 벡터에 저장하고 결과물을 아래와 같이 출력되게 보여주세요

코드

```
txt <- read_file('http://www.gutenberg.org/cache/epub/1112/pg1112.txt')
Encoding(txt) <- "UTF-8"

script<-unlist(strsplit(txt,'ACT'));script
script1<-unlist(strsplit(script,'\r\n'));script1
script2<-grep('Scene',script1);script2
script3<-script2+2
script1[script3]
```

결과

```
> script1[script3]
[1] "Verona. A public place."
[3] "Capulet's house."
[5] "Capulet's house."
[7] "Capulet's orchard."
[9] "A street."
[11] "Friar Laurence's cell."
[13] "Capulet's orchard."
[15] "Capulet's house."
[17] "Friar Laurence's cell."
[19] "Juliet's chamber."
[21] "Juliet's chamber."
[23] "Verona. Friar Laurence's cell."

"A Street."
"A Street."
"A lane by the wall of Capulet's orchard."
"Friar Laurence's cell."
"Capulet's orchard."
"A public place."
"Friar Laurence's cell."
"Capulet's orchard."
"Capulet's house."
"Capulet's house."
"Capulet's house."
"Mantua. A street."
"Verona. A churchyard; in it the monument of the Capulets."
```

문제3)

코드

```
article = read_file('article.txt')
q<-unlist(strsplit(article,['\r\n']));q
w<-grep('s',q);w
r<-q[w]
r[1:4]
```

결과

```
[523] "\"It\"      \"is\"      \"a\"      \"shame\"    \"we\"      \"are\"      \"not\"      \"close,\"\" \"their\"
[532] \"manager,\"  \"Mauricio\" \"Pochettino,\" \"said.\"  \"\"we\"     \"were\"     \"but\"     \"in\"      \"the\"
[541] \"last\"      \"few\"      \"weeks\"     \"we\"     \"find\"     \"our\"     \"reality,\" \"our\"     \"place.\"
[550] \"we\"        \"are\"      \"disappointed\" \"but\"    \"knowing\"  \"realistically\" \"we\"      \"must\"     \"be\"
[559] \"happy.\"    \"\"        \"\"        \"\"        \"\"        \"\"        \"\"        \"\"        \"\"
[568] \"\"        \"\"        \"\"        \"\"we\"    \"have\"     \"a\"        \"mini\"    \"league\"   \"now,\"
[577] \"seven\"     \"games.\"  \"It's\"     \"in\"     \"our\"     \"hands\"    \"to\"      \"be\"      \"in\"
[586] \"the\"      \"top\"     \"four\"     \"and\"    \"the\"     \"champions\" \"League\"  \"next\"    \"season.\"
[595] \"\"        \"\"        \"\"        \"\"        \"\"        \"\"        \"\"        \"\"        \"\"
[604] \"\"        \"\"        \"COPYRIGHT\" \"-\"      \"THE\"     \"TIMES,\"  \"LONDON\"
> w<-grep('s',q);w
[1] 2 71 337 380 609
> r<-q[w]
> r[1:4]
[1] \"SALAH\" \"salah\" \"Sissoko\" \"Son\"
```

문제4) 아래 birthday란 변수에 본인의 생년월일을 문자열로 지정하고 본인이 태어난 지 몇 일째 되는지 확인하는 코드를 작성하고 태어난 날의 요일은 무슨 요일인지 확인하는 코드를 작성 후 결과물을 아래와 같이 출력되게 보여주세요

코드

```
birthday = '20001229'
recent_time<-Sys.Date()
birthday1<-as.character(birthday)
birthday2<-as.Date(birthday1,'%Y%m%d')
difftime(recent_time,birthday2,tz=\"EST\")
weekdays(birthday2)
```

결과

```
> birthday = '20001229'
> recent_time<-Sys.Date()
> birthday1<-as.character(birthday)
> birthday2<-as.Date(birthday1,'%Y%m%d')
> difftime(recent_time,birthday2,tz="EST")
Time difference of 6673 days
> weekdays(birthday2)
[1] "금요일"
```